

# The Self

Nigel Warburton







## Book 1

# The Self

Nigel Warburton

This publication forms part of the Open University module A222 *Exploring philosophy*. Details of this and other Open University modules can be obtained from the Student Registration and Enquiry Service, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)845 300 60 90; email [general-enquiries@open.ac.uk](mailto:general-enquiries@open.ac.uk)).

Alternatively, you may visit the Open University website at [www.open.ac.uk](http://www.open.ac.uk) where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

To purchase a selection of Open University study materials visit [www.ouw.co.uk](http://www.ouw.co.uk), or contact Open University Worldwide, Walton Hall, Milton Keynes MK7 6AA, United Kingdom for a brochure (tel. +44 (0)1908 858793; fax +44 (0)1908 858787; email [ouw-customer-services@open.ac.uk](mailto:ouw-customer-services@open.ac.uk)).

The Open University Walton Hall, Milton Keynes MK7 6AA

First published 2011

Copyright © 2011 The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilised in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS (website [www.cla.co.uk](http://www.cla.co.uk)).

Open University study materials may also be made available in electronic formats for use by students of the University. All rights, including copyright and related rights and database rights, in electronic study materials and their contents are owned by or licensed to The Open University, or otherwise used by The Open University as permitted by applicable law.

In using electronic study materials and their contents you agree that your use will be solely for the purposes of following an Open University course of study or otherwise as licensed by The Open University or its assigns.

Except as permitted above you undertake not to copy, store in any medium (including electronic storage or use in a website), distribute, transmit or retransmit, broadcast, modify or show in public such electronic materials in whole or in part without the prior written consent of The Open University or in accordance with the Copyright, Designs and Patents Act 1988.

Edited and designed by The Open University.

Printed in the United Kingdom by Page Bros, Norwich.



ISBN 978 1 8487 3309 1

1.1

# Contents

Preface	v
1 What am I?	1
2 Locke on persons	27
3 The self as a fiction	53
4 The self and the future	77
Readings	103
Glossary	125
Acknowledgements	128
Index	129



## Preface

This is the first of six books in A222 *Exploring philosophy*. It is somewhat shorter than the other books, in order to allow you to get used to the demands of studying the subject, possibly for the first time, and to give you longer to get to grips with the texts from seventeenth- and eighteenth-century philosophers which form part of this month's work.

By the end of this book you should have a good overview of the key writers on the topic of personal identity. You should also have had practice reading original philosophical texts and have thought critically about some of the ideas expressed. In addition, the topic of thought experiments and their importance in philosophy crops up in several places.

You will start by reading the work of the seventeenth-century philosopher John Locke. His ideas about personal identity are the starting point for most modern discussions of the topic. We then move on to a classic extract from the writings of David Hume, in which he questions whether introspection gives us any evidence for an enduring self. The final chapter concentrates on the contemporary philosopher Derek Parfit's controversial views about the self.

The relevant readings can be found in the set book, John Cottingham's *Western Philosophy*, and at the back of this book. You'll be guided through these in the main text. Don't be disheartened if you find some of the reading difficult at first. It is probably a good sign, as anyone who finds it really easy is probably missing something. There is also a glossary at the back of the book, which explains some key terms. These terms are emboldened at their first (or at least an early) use in the text.

The audio recordings that accompany this book are an essential and integral part of the teaching. They include interviews with a range of eminent philosophers. At various points you will be directed towards them. Ideally, you should listen to the audio recordings at the point suggested in this text, but you may want to listen to them all quite early on as well to get an overview of where you are going. If it is impractical for you to listen at the points recommended in the text, you may want to print out transcripts (which are available on the *Exploring philosophy* website) to allow you to study the material in the appropriate order. You can then consolidate this by listening to the recordings later.

It is important to keep in mind that studying philosophy should not be a passive experience. You will be encouraged to do exercises, discuss ideas, write, listen actively and think for yourself. Avoid the temptation simply to absorb what you read without really engaging with it.

Philosophy is not a spectator sport. In order to study it, you have to do what the great philosophers of the past did: think critically about some of the most profound topics that we can ask ourselves.

I hope you enjoy studying this book. Philosophy can be a demanding subject, particularly when you first begin studying it, but it is also extremely rewarding.

# Chapter 1

## What am I?



# Contents

Aims	5
Materials you will need	6
1.1 Introduction	7
1.2 The Ship of Theseus	11
Thought experiments in philosophy	14
1.3 Personal identity	17
Three thought experiments	19
Summary	25
References	26



# Aims

By the end of this chapter, you should:

- have an overview of the philosophical problem of personal identity
- understand what a thought experiment is and why philosophers use them
- appreciate some of the main features of John Locke's account of personal identity.

## Materials you will need

You will need to listen to the following audio recordings (available on the *Exploring philosophy* website):

- Thought experiments
- Personal identity.

You will also be directed to the website for an end-of-chapter quiz.

## 1.1 Introduction



Figure 1.1 Self-portraits by Rembrandt, painted in 1629, c.1637, 1657 and c.1668–69. From left to right: *Head of a Young Man or Self Portrait*, 1629, oil on panel (Alte Pinakothek, Munich; photo: © Alte Pinakothek, Munich/Giraudon/The Bridgeman Art Library); *Self Portrait*, c.1637, oil on oak panel (Wallace Collection, London; photo: © Wallace Collection/The Bridgeman Art Library); *Self Portrait*, c.1657 (Kunsthistorisches Museum, Vienna; photo: © Austrian Archive/Scala, Florence); *Self Portrait*, c.1668–69 (Wallraf-Richartz Museum, Cologne; photo: © Wallraf-Richartz Museum, Cologne/The Bridgeman Art Library).

Over a period of about forty years, the great Dutch artist Rembrandt van Rijn (1606–1669) drew, etched or painted himself at least seventy-five times. Despite the many poses and roles he adopted, and the numerous and poignant physical changes his body underwent in this time, we can see that the young boy depicted in the early work is in some sense the same person depicted as an old man in the late self-portraits. Perhaps it is something about the eyes – we feel it is the same person looking out at us at different stages of his life. As he painted or etched these remarkable self-portraits he would have been looking carefully at his mirror image, but he, Rembrandt, the same Rembrandt, also seems to be looking out at us from all of them.

One source of the paintings' power is this: they show the continuity and change that is an essential part of the human condition. We can now see the whole series, and also we recognise the adult in the boy, and the old man in the middle-aged one. There is a sense in which Rembrandt is the same throughout while undergoing changes to his body, and to his fortunes. Anyone who has seen the astounding Granada documentary television series *Seven Up!*, which follows the lives of fourteen children who were seven years old in 1964, revisiting them every seven years throughout their lives, will appreciate this idea of continuity of some aspects of character despite physical change and

some change of outlook. Indeed, the title for the show comes from a Jesuit saying: 'Give me a child until he is seven and I will give you the man.' Most of us will age very visibly, and our characters will change subtly, or, perhaps, more radically – we might become more cynical and jaded than we were in our youth, or possibly more charitable and less ambitious for ourselves – but we will remain continuous with the people we were in childhood up to the point when we die, when the changes are more extreme. Or at least that's how it feels.

Some people believe that they will continue to live – though not necessarily in a physical form – after their deaths. I'm not one of these. For those who do hope for post-mortem existence, though, the question 'What will survive and how will it relate to what I was when I was alive?' is a poignant one. In more religious ages, the question of whether we might be resurrected bodily after death, and where that left people whose corpses had been destroyed or dispersed, was of utmost importance.

Belief in some kind of afterlife is extremely widespread, and there is still a serious philosophical question about what could possibly survive and in what sense this would or could be the same person as was alive on earth. If we are essentially physical beings requiring physical continuity of a particular body for continued existence, then some kind of Day of Judgement resurrection of everyone's body seems to be required if we are to be the same people in an afterlife. If what matters in questions of continuing personal identity is a variety of psychological rather than physical continuity, the particular body (or other form) you end up with may not be so important. If to be the same person in the future you just have to be able to remember your own past, for example, whether the body you have is historically continuous with the one you had in life, or whether you even have a body at all, may not matter so much. There are further options, too – many Hindus, some Buddhists and even some non-religious people believe that our souls will be reincarnated in different bodies on earth after death even though we will have no recollection of having lived before. In the Hindu sacred scripture, the *Bhagavad Gita*, Krishna declares:

Just as a man casting off worn-out clothes takes up others that are new, so the embodied self, casting off its worn out bodies, goes to other, new ones.

(Johnson (trans.), 1994, p. 9)

Seventeenth-century concerns about the Day of Judgement are discussed in Chapter 2.



Figure 1.2 Sir Stanley Spencer, *The Resurrection, Cookham*, 1924–27, 274 x 549 cm, oil on canvas. Tate Britain, N04239. Presented by Lord Duveen, 1927. Photo: © Tate, London, 2010.

This implies that the self is separable from the particular body, that it can move from body to body, and also that memory of past bodies and lives is not a prerequisite for continued existence as a self.

But even for the non-religious there is an interesting question (that may one day become a practical one), about whether it might be possible to carry on living in a different body from the one you now have. If a scientist of the future could transfer all my memories from my brain to some other donor brain, would I carry on living in that other brain? Or would that just be another way of dying? Am I the software that happens to be instantiated in the particular hardware of my brain now, but which could, in principle, be run in a different physical system? Or am I fundamentally this particular animal, a functioning body with conscious experience that could not exist apart from it?

It might seem obvious that what makes you the same person from day to day despite small physical changes is bodily continuity: whatever else you are, you are a member of the species *Homo sapiens*, have a limited lifespan, and will necessarily age in more or less predictable ways until death comes. You might think that what makes you the same person over time has nothing to do with your character, but is rather a matter of your physiology. Yet is it simply the physical continuity of your body that makes you in an important sense the same person that you were

ten years ago, though, we hope, ten years wiser, and more mature? Is my body like a ship that gets a plank replaced one at a time until every piece has been changed? Even if it is, it may not be straightforward to say that I am the same person that I was.

In this book, we'll be exploring a range of related questions about personal identity, focusing on the questions 'What makes a person the same person now as they were at earlier times, and why does it matter?' We'll be reading extracts from the works of the seventeenth- and eighteenth-century philosophers John Locke and David Hume, as well as considering the approach a contemporary philosopher, Derek Parfit, takes to these questions. Locke is generally credited with introducing serious discussion of the nature of a person over time in to philosophy in the second edition of his *Essay concerning Human Understanding* (published in 1694), though some classical philosophers had discussed similar questions much earlier.

Before getting to grips with Locke's ideas on this topic, we are going to look at some more general issues about identity over time for non-living things. We will also be examining the strategy of using a type of hypothetical example known as a 'thought experiment' – an important technique in philosophy and one that is much used in the discussion of personal identity.

## 1.2 The Ship of Theseus

The first-century Greek historian Plutarch described the case of the Ship of Theseus. This was the ship in which Theseus travelled to Crete and back where he slew the Minotaur. After Theseus' homecoming, his ship was kept in Athens. As parts of it rotted, they were carefully replaced with new wood. Eventually, many years later, there were no timbers left from the original vessel – everything was new. Was the ship at the end of this process the same ship as the one in which Theseus sailed back to Athens? Philosophers at the time were divided on the issue: some said 'Yes', others 'No', and those on each side could give reasons for their answer. Here is Plutarch's description:

The ship wherein Theseus and the youth of Athens returned had thirty oars, and was preserved by the Athenians down even to the time of Demetrius Phalereus, for they took away the old planks as they decayed, putting in new and stronger timber in their place, insomuch that this ship became a standing example among the philosophers, for the logical question of things that grow; one side holding that the ship remained the same, and the other contending that it was not the same.

Plutarch, 2001, p. 21

### Activity

Note down brief answers to the two questions below before reading the discussion that follows. If you don't have a pen and paper handy, at least think for a moment about each of the questions before looking at the discussion.

- 1 In what sense was the reconstructed ship the same one in which Theseus sailed to Crete and back?
- 2 In what sense was the reconstructed ship a different ship from the one Theseus sailed?

### Discussion

- 1 The ship was repaired a plank at a time, so is continuous with the one that Theseus sailed. It also functions in the same way. It's *numerically* the same ship, just as the tea clipper the *Cutty Sark* in Greenwich, which was badly damaged by fire in 2007, remains the *Cutty Sark* after extensive reconstruction. Even if no original parts remained (unlike the *Cutty Sark* case, where many original parts were salvaged and integrated), there is a sense in which the ship of

Activities, like this one, crop up throughout *Exploring philosophy*. You may be tempted to skip them and simply read on to the discussion sections, but this is not a good idea! Activities are intended to help you understand the key issues, and to approach the teaching materials in an active way, rather than letting them simply wash over you. You are strongly advised to attempt them all as they occur in the text.

Theseus in the time of Demetrius Phalerius is the same ship as the one Theseus sailed since it arose directly from repairs to the old one. This makes it very different from a replica made somewhere else that would not have this continuity. The replica might be *qualitatively* more similar to the original than the reconstructed ship, i.e. it might more closely resemble it, be a superb copy of it, but that wouldn't make it *the same ship*, only *the same type of ship*.

- 2 The ship in the time of Demetrius Phalerius is *qualitatively* different from the one Theseus sailed – the planks are newer. Theseus didn't walk on the wood from which the deck was made, for instance. These differences might lead some people to declare that this wasn't the same ship at all.

---

As Plutarch noted, philosophers had different opinions as to whether the ship of Theseus had been preserved in the time of Demetrius Phalerius or whether it was a different ship, a replica of the original. What this case reveals is that the phrase 'the same as' can mean different things in different contexts. The seventeenth-century philosopher John Locke presented a similar conundrum with his example of a sock that is repeatedly patched up until no wool of the original sock remains. Is it then the same sock that it was formerly? The identity of persons over time is a special case of a more general philosophical issue about the continuing identity of objects that change over time.

### Qualitative and numerical identity

Words or phrases in bold can be found in the glossary at the end of the book.

If two things are **qualitatively identical**, that means they are of the same type. If they are **numerically identical**, that means they are the same thing. So, for example, identical twins have qualitatively identical chromosomes, but are still numerically distinct: that is, there are two of them. The Morning Star and the Evening Star are numerically identical: that is to say, they are one and the same thing (The Morning Star and the Evening Star are two names used to refer to the planet Venus).

For two things to be strictly identical, everything that is true of one is true of the other. This is easiest to see in mathematics. When we say that the sum of two plus two is identical to the sum of three plus one, we mean that there is no difference whatsoever in any respect between

the two totals: whatever you can say about one is true of the other. Yet when we talk about the identity of living things, such as trees and people, we don't employ this strict sense. When someone asks in the autumn 'Is this the tree we sat under in the summer?', the fact that the leaves have now fallen does not prevent its being the same tree. What we mean by 'the same tree' here is that it is the same living plant despite changes that it has undergone. It is numerically identical with the tree we sat under though qualitatively very different. In other words, when it comes to the identity of living things we use different criteria from those for strict logical identity in mathematics or geometry.



Figure 1.3 The *Cutty Sark* undergoing restoration. In what sense is the reconstructed ship still the *Cutty Sark*? Photo courtesy of The Cutty Sark Trust.

Similarly we might want to say that what makes us the same people we were ten years ago (though our bodies have changed significantly, many cells have been replaced, our hair and teeth might have fallen out, our best time for running a 10 km race is much slower, and so on) is that each of us is the same biological organism that we were, because the change, like the change in the ship, has been piecemeal, and yet we recognise that in a different sense we have changed a great deal.

Thomas Hobbes (1588–1679), a major English philosopher most famous for his book *Leviathan* (1651) in which he used the phrase ‘solitary, poor, nasty, brutish and short’ to describe life outside society in what he called ‘the state of nature’.

Julian Baggini refers to this case in the audio recording ‘Thought experiments’ which you will listen to shortly.

The audio recording ‘Thought experiments’ lasts about twelve minutes – but you will need to listen to it more than once.

In the seventeenth century, the great English philosopher Thomas Hobbes – in his book *De Corpore* (1655) – suggested a variant of the Ship of Theseus case, one that complicates matters. Imagine that when the planks on the ship are replaced someone picks up the old pieces and eventually manages to knock together a complete ship. Now we’d have a situation where there are two candidates: the ship that has been reconstructed using new timber, and the one that has been built from the scavenged scrap. Which, if either, of these is truly the Ship of Theseus? This is not an easy question to answer.

## Thought experiments in philosophy

The case of the Ship of Theseus that we have been looking at is an example of a **thought experiment**. These are widely used in philosophy and you will come across many of them in *Exploring philosophy*. They consist of a hypothetical situation in which variables are carefully controlled and an invitation to consider what you would say or do in that situation.

### Activity

Now listen to the audio recording ‘Thought experiments’ twice. In it, I discuss thought experiments with Julian Baggini, author of *The Pig That Wants to Be Eaten* (2006). The interviewer Winifred Robinson introduces the recording and asks questions from the point of view of someone studying *Exploring philosophy*. The first time you listen, play it straight through to the end so that you get an overview of the topic; the second time, make brief notes on the questions below before going on to read the discussion.

- 1 What are thought experiments generally intended to reveal?
- 2 (a) What do thought experiments have in common with scientific experiments?  
(b) How do scientific experiments and thought experiments differ?
- 3 What is Baggini’s response to Hobbes’s version of the Ship of Theseus thought experiment (in which a ship is rebuilt from the discarded planks)?
- 4 Towards the end of the interview, Baggini suggests that real examples may in some situations be preferable to thought experiments. Give your own real-life example of something that changes over time and raises similar issues to the Ship of Theseus about identity.

## Discussion

- 1 Thought experiments are devices for revealing our intuitions. By describing hypothetical situations, philosophers isolate what is important in an issue and bring our pre-reflective judgements into the open for analysis. They are often highly imaginative, bringing in science fiction elements. Their point is not to entertain, however, but to help us get clear about what we believe. When we bring these intuitions into the light we can see whether or not they are simply prejudices.
- 2 (a) Like scientific experiments, thought experiments isolate variables, allowing us to focus on the key points at issue. They also manipulate these variables. So, for example, in the trolley problem thought experiments I describe in the audio recording there are two related cases: the case of the runaway trolley that I can redirect and the case of the runaway trolley that I can stop by pushing a large person off a bridge. The point here is that both look to be cases of sacrificing one person to save many. But my intuitions about what I should do are different in the two cases. This reveals that questions about how that individual comes to die and how I bring that about may be relevant. By manipulating the variable, in this case describing two related scenarios, philosophers hope to be able to isolate the important points at stake.  
(b) We should not push the analogy between thought experiments and scientific experiments too far. Thought experiments involve thinking about what we might do in hypothetical situations; scientific experiments focus on what actually happens in real circumstances.
- 3 Baggini points out that there isn't a correct answer to the question 'Which of these ships is the Ship of Theseus?' It all depends on why you are asking the question. From a legal point of view it is the repaired one; from a forensic point of view, if, for example, we are investigating a murder at the time of Theseus, it would be the one made from the old parts that would be most relevant.
- 4 You might have a bicycle that has been repaired so many times that there is an issue of whether it is the same bicycle that you originally bought, or thought of a church or other building that has been repeatedly rebuilt, or something similar. There are several pop bands whose personnel have totally changed since the band was first formed. If the original members got together again, which would be the real band? These are just a few examples of the kind of case you might have found. In most real examples, it is unlikely that the perfect

case of every part being replaced will have been met; whereas it was easy for Hobbes to imagine this situation to draw attention to the complexity of our decisions about the identity of objects that change over time. That is one advantage thought experiments can have over the slightly messier nature of real examples: by eliminating distracting factors it is possible to zoom in on what is more relevant to the topic under discussion.

---

You will encounter numerous thought experiments in the readings for this book. They are invitations to make you think issues through for yourself rather than simply to absorb material passively. This is a typical characteristic of the best philosophical writing: it is designed to make you think as you read it; and it requires you to engage actively as you mull over issues raised. Studying philosophy involves learning to philosophise rather than simply learning what other philosophers have thought. This is one reason why reading philosophical writing can seem tough if you are not used to it.

## 1.3 Personal identity

We are now going to return to the question of personal identity; the question of what, if anything, makes an individual the same person over time. This question isn't just relevant to what may or may not happen to us after death. Nor is it merely a logical puzzle about how we use the words. There are serious issues about holding people responsible for their actions that are tied to personal identity. Where there is a question of moral responsibility, where we want to know whom we should praise or blame for a past action, we seek out *the person who performed the good action or the crime*. If someone committed a murder years ago, we usually feel that, despite ten years of ageing, that person is still the murderer. We typically justify punishing that person when they are caught on the grounds that they are still the culprit, still the person morally and physically responsible for their past actions, despite the psychological and physical changes that have occurred in the interim. The question of what we mean by 'the same person' here, given this undeniable change over time, is, then, central to our notion of moral responsibility. It is a question that has vexed philosophers since the time of the ancient Greeks.

When a former Auschwitz guard is tracked down sixty years after committing atrocities, and he can't remember anything about it, perhaps as a result of Alzheimer's disease, there is a question as to whether the old man in front of us is the same person who acted so viciously in his youth. There is a straightforward question about identity that sometimes arises in such cases: is this the right human being we have in the dock? Could this be a case of mistaken identity? Perhaps the person charged is completely innocent and is just unfortunate enough to look a bit like the real culprit.

There is, however, another important question, a moral one. That is the question of whether or not individuals are now morally responsible for what they did so long ago. Assuming that their actions at the time were voluntary (i.e. they weren't forced to commit atrocities), and setting aside questions about whether or not they are feigning forgetfulness, is genuine amnesia an adequate excuse in such cases? If you sincerely forgot crimes you committed twenty years ago, should you be punished for those past actions? This question is further complicated by another philosophical question, namely, 'What is the point or justification of punishment?'

Though far-fetched today, it may in the future be possible to transfer memories from a dying person to an artificially created brain – or perhaps, to one donated from someone who has died. Would the person with the newly created brain with the now-dead individual's memories be the same person as the one whose body died? Whether or not this remains in the realm of science fiction, it is still an interesting question to ask for what it reveals about our understanding of what a person essentially is.

### personal identity

The term 'personal identity' is much used by philosophers when discussing what, if anything, makes a person the same person despite changes over time. This should not be confused with the political sense in which 'identity' is sometimes employed, as when people refer to an individual's identity in terms of their cultural, national and ethnic origins, sexual preferences, and so on. In this book, I'm focusing exclusively on philosophical questions about the self over time, rather than on an individual's identity in this other political or sociological sense.

### Activity

The audio recording 'Personal identity' lasts about ten minutes – but you will need to listen to it more than once.

Listen to the audio recording 'Personal identity', in which I interview the philosopher A.C. Grayling. As with all the audio recordings, you may want to listen to it several times, and this is what we recommend. Listening to philosophical discussion is a skill in itself that takes practice to perfect. This recording is intended to give you an overview of the main topics discussed in the rest of this book. Grayling outlines the approach taken by John Locke, David Hume and Derek Parfit as well as explaining what the basic problem of personal identity is and why it is important. As with the previous activity, listen to the whole recording at least once to get an overview before replaying it and jotting down brief answers to the following questions (all of which relate to Locke's writing on personal identity as discussed in the first five minutes of the recording).

- 1 Why, according to Grayling, does the problem of personal identity pose problems for philosophers?
- 2 Why, according to Grayling, is Locke not content to say that persons are simply like acorns that turn into oak trees, i.e. organisms that change over time?

- 3 Would Locke hold a former concentration camp guard who had forgotten what he had done in the camp morally responsible for his actions there?

## Discussion

The point of listening to this audio recording at this stage is to get an overview of the problem of personal identity and how the main philosophers studied in this book – John Locke, David Hume and Derek Parfit – deal with it. Don't worry if you can't follow every point at the moment. When you re-listen to this audio recording having studied the whole book, much more should make sense to you. For the moment I want to concentrate on Grayling's summary of Locke's views. We will be looking at an extract from Locke's writing in Chapter 2; Grayling's overview of Locke should give you some pointers to the most important features of Locke's account.

Here are my answers to the questions.

- 1 Personal identity raises issues that are of crucial importance in ethics, the philosophy of law, and in the area of the individual and society. The question of what makes a person the same person over time relates directly to questions about responsibility and rights.
- 2 Bodily continuity explains the identity of plants and animals over time. But for Locke, the notion of a 'person' is not the concept of a physical thing. It is a 'forensic' concept, that is, a concept that relates to questions of moral responsibility for actions. The implication is that these can't be reduced to questions of bodily continuity.
- 3 Locke believed that moral responsibility requires memory of the actions for which we are held responsible. So if the guard had genuinely forgotten what he had done, then, on Locke's analysis, he should not be held morally responsible for these actions.

---

## Three thought experiments

In this section, we will look at three scenarios, all of which are based on thought experiments used by John Locke. After each scenario is described, there is a short activity. Write down a brief answer for each, in which you try to give your own opinion. If you don't have writing materials to hand, at least think through the questions.

## Who was John Locke?



Figure 1.4 Unidentified after H. Garnier, *John Locke*, lithograph. Library of Congress Prints and Photographs Division, Washington, DC. Library of Congress.

Locke (1632–1704) was a British philosopher and political theorist. He is best known for his idea, expressed in his *Essay concerning Human Understanding* (first published in 1690), that the child at birth is like a ‘*tabula rasa*’ (Latin for ‘blank slate’) and that experience teaches us everything we know: there are no innate ideas. He saw his role as that of an ‘under-labourer’, helping to clarify the concepts that scientists used (i.e. clearing away the intellectual debris from confusions caused by imprecise ideas). He also made significant contributions to political theory and was a passionate defender of religious toleration. Although principally a philosopher, he did have an active interest in science, and even performed a life-saving medical operation on his patron.

### Scenario 1: The prince and the cobbler

A prince wakes up one morning and looks in the mirror. To his surprise and consternation, instead of the dashingly handsome and well-groomed, blue-blooded and highly moisturised specimen he is used to encountering, he sees the rugged pock-marked face of a poor cobbler: a shoemaker whose skin has been discoloured by the tanning dye. Where once he saw the elegant profile of one who was selectively bred from

royalty, now he sees what he thinks of as the crude features of a common worker. He can remember going to bed the night before, throwing his beautiful clothes on the floor for his man-in-waiting to pick up and fold, putting on silk pyjamas and turning out the bedside light. When he shuts his eyes he is in every sense still the prince; it is only externally he has changed. Like Gregor Samsa – the character in Kafka's short story 'Metamorphosis', who wakes up to discover that he has turned into an insect – something terrible has happened to him. He is still himself in one sense, but not himself in another.

At precisely the same time, a cobbler wakes up, scratches himself as usual, and is disconcerted to see that, in place of his strong worker's hands with split fingernails, by some strange magic he has acquired the long elegant fingers and carefully manicured cuticles of a prince. He looks in the mirror and finds himself staring at the face of the prince. His whole body has been transformed somehow into the prince's body – or, rather, his memories and mind have somehow migrated there.

---

### Activity

Imagine that several days before these strange events the prince had just murdered his elder brother in order to accelerate his ascendancy to the throne. Your task is to decide which of these two characters, the one with the prince's body and the cobbler's memory, or the one with the prince's memory and the cobbler's body, should be charged with the murder and punished? In other words, which – if either – of the two do you believe deserves punishment?

---

### Scenario 2: The day-man and the night-man

A man called Socrates is not quite what he seems. He appears to be just one person, but in fact within the one body there are two distinct people. By day the day-man goes about his business, thinks his thoughts, and so on; but by night a completely distinct consciousness inhabits this body. The night-man has a very different personality from the day-man. The day-man is a thoughtful philosopher. The night-man hates philosophy: all he cares about is golf. Neither knows of the other's existence. Neither has access to the other's thoughts.

This is not quite as far-fetched as it initially sounds. Although rare there are genuine cases of multiple personality where more than one distinct consciousness seems to occupy the same body and where there is no

communication of thought between the two consciousnesses. Also, on occasion, neurosurgeons sever the *corpus callosum*, the bundle of nerve connections between the two hemispheres of the brain, and this drastic measure, which is occasionally used to treat severe epilepsy, can lead to the creation of what appear to be distinct and independent consciousnesses within the same individual's skull. A person in this condition may select a dress from the wardrobe with one hand while the other one tries to put it back.

---

### Activity

If the night-man loses his temper on the all-night golf drive and hits someone with his club, do you think it would be right to punish him by day (i.e. to punish the day-man)?

---

### Scenario 3: The drunk and the sober man

If you find the previous two thought experiments too far-fetched to take seriously, then consider the case of a drunk. This will probably require less imagination. This man goes out one evening in New York. After downing most of a bottle of bourbon, he staggers out of the bar and becomes involved in a street brawl, breaking his opponent's nose with a lucky left hook. By a near miracle he manages to make his way back to his hotel room and falls fast asleep in a drunken stupor. When he wakes up in the morning he can remember nothing whatsoever about what happened after he left the bar. However hard he tries, he just cannot recall those last hours.

---

### Activity

If, on the morning after, the drunk genuinely couldn't remember breaking the other man's nose, nor the events that led up to this, would it, in your opinion, be morally right to punish him for what happened?

---

All three of these cases raise issues about the importance of one's memory of events when describing moral responsibility for action.

- With the prince and the cobbler, most people's response is that, in such a weird case, it would seem right to punish the person with a cobbler's body but the prince's consciousness on the grounds that the man with the cobbler's memories knows nothing about the event; whereas the man with the prince's memories knows all about

the murder, the motives, the method, and remembers in detail precisely what happened.

- Similarly, in the case of the day-man and night-man it seems wrong to many of us to punish the day-man for the activities of a distinct personality that happens to inhabit the same body.
- Although in most real cases we might be sceptical about the suggestion that someone had completely forgotten their crime, it might seem to be a mitigating circumstance that the drunk had no knowledge whatsoever of what he had done. This case is trickier in many ways than the other two. Many of us would hold this person responsible for getting so drunk that he wouldn't be able to remember what he had done. You might feel that forgetting having committed a crime doesn't mean that it wasn't you who committed it – who else did it, after all?

A further aspect of this, which you might have considered, is that the point of punishment might not simply be to inflict the punishment on the right person, but might also serve to communicate to other potential criminals that there is a risk of punishment for certain behaviour. If there is a deterrent aspect to punishment, it might not matter so much whether or not the person punished remembers their crime.

Your answers to these questions may well be different from mine. But I hope you can understand why some people feel that continuing memory is at least relevant to questions about moral responsibility for our actions in at least some cases, why it might be wrong to punish or praise someone for something that they did, but which they genuinely couldn't remember at all – though in real-life cases we might be sceptical about convenient 'forgetting'. This is the position Locke took. In his versions of the three scenarios discussed above, memory is the criterion of personal identity – and thus of moral responsibility – not bodily continuity. So, for him, in Scenario 1, the individual with the memory of the murder would be the one who should be punished. In Scenario 2, the punishment would only be merited when the individual remembers the actions. And in Scenario 3, if the sober man genuinely couldn't remember what he did when drunk, then he would not be morally responsible for those actions.

Make sure that you have thoroughly understood this chapter before moving on to the next one. You may find it useful at this stage to listen again to the audio interviews with Julian Baggini and A.C. Grayling to

You will consider theories of punishment further in Book 3, Chapter 2.

You can find lists of 'Optional further material' on the *Exploring philosophy* website.

consolidate what you have learned so far. If you have sufficient time, I recommend browsing through the list of suggestions in the 'Optional further material' for this chapter. Among other things, this includes highly relevant entries from the *Routledge Encyclopedia of Philosophy* and the *Oxford Dictionary of Philosophy*, both available through the Open University Library website.

---

### Activity

Listen to the remaining audio recordings for this book (that is, all the recordings that you haven't yet listened to). You needn't do this in a single sitting. Don't worry if you don't follow every aspect of all the discussions.

### Discussion

The point of this activity is to give you an overview of this book before you get immersed in the details of the readings. The discussions will become clearer when you have studied the accompanying texts. You will be instructed to listen to all these recordings again individually at key points in the next three chapters.

---

# Summary

The key points we have covered in this chapter are:

- There is an important philosophical question about what a person is that has implications in ethics and the philosophy of law.
- Thought experiments are imaginary cases in which variables are controlled to reveal intuitions. They are much used in philosophy.
- Seventeenth-century philosopher John Locke believed that what makes someone the same person over time was self-consciousness and continuity of memory rather than bodily continuity.

In this chapter, I have introduced the technique of using thought experiments in philosophy. You have also practised the skill of listening to discussions of philosophical topics. In Chapter 2 we will be looking more closely at Locke's writing on personal identity and thinking more generally about reading philosophical texts from the past.

---

## Activity

On the *Exploring philosophy* website, you will find quizzes to help you revise each week's work. Have a look at the quiz for this chapter now before moving on.

---

## References

Johnson, W.J. (trans.) (1994) *The Bhagavad Gita*, Oxford, Oxford World's Classics.

Locke, J. (1975 [1690]) *An Essay concerning Human Understanding* (ed. P.H. Nidditch), Oxford, Clarendon.

Plutarch (2001) *Plutarch's Lives*, ed. Hugh Arthur Clough, trans. John Dryden, New York, Modern Library.

# Chapter 2

## Locke on persons



# Contents

Aims	31
Materials you will need	32
2.1 Introduction	33
2.2 Preparing to read Locke	35
Locke's terminology	35
2.3 Focusing in on Locke's <i>Essay</i>	39
2.4 Engaging with Locke	43
Thomas Reid's criticism of Locke	43
J.L. Mackie's reading of Locke	46
2.5 Reading historical texts	48
Summary	51
References	52



# Aims

By the end of this chapter, you should:

- have understood the main features of John Locke's account of personal identity
- have a good understanding of the key five paragraphs from the Locke reading
- have considered Thomas Reid's counter-example to Locke's memory criterion of personal identity
- be aware of some of the ways that philosophers use historical texts.

## Materials you will need

In this chapter you will need to refer to the following reading in the set book:

- Part V, Reading 1, 'The Self and Consciousness: John Locke, *Essay concerning Human Understanding*' (focusing on five paragraphs).

You will need the following additional readings, which can be found at the end of this book:

- Reading 1: Locke on personal identity (Bennett's paraphrase)
- Reading 2: Mackie on the unity of consciousness.

You will need to listen to the following audio recordings (available on the *Exploring philosophy* website):

- Locke on persons
- Reading historical texts.

You will also be directed to the website for an end-of-chapter quiz.

## 2.1 Introduction

This chapter has two parts. The first part concentrates on reading John Locke's writing on personal identity. The final section will look at more general questions about reading philosophers of the past.

The three scenarios we considered at the end of Chapter 1 are all my elaborations of thought experiments that John Locke used in Chapter 27 of his book *Essay concerning Human Understanding* (second and later editions). The passage that forms the reading 'The Self and Consciousness' in the set book (Part V, Reading 1) includes all three, as well as several more, thought experiments. This passage is the most famous discussion of what a person is in the history of philosophy and is usually taken as the starting point for all further debate about the topic, though in fact some ancient Greek philosophers had discussed this and related questions almost 2000 years earlier. What Locke did was to raise some of the key philosophical issues that arise when thinking about what a person is and how notions of personal identity over time relate to questions of moral responsibility for past actions. He also drew influential conclusions about what he thought was important for continued identity as a person.

The rest of this week's work will be taken up with trying to make sense of part of this short reading by Locke – after getting a sense of the whole piece, we're going to focus on just five paragraphs. This is not an easy read by any means, and so I will do everything I can to help you make sense of the main ideas within it. Don't be discouraged by the initial impenetrability of the text: it should be much clearer by the time you have worked through the teaching materials. But for most readers this piece will present quite a challenge. Philosophers still argue over the precise meaning of some passages in the reading. Despite its complexity, though, the main thrust of it is that what makes a person now the same person as a person in the past is continuity of memory. Try to remember this main idea when you are immersed in the detail of the arguments and examples Locke uses – it is easy when dealing with a dense piece of writing like this to fail to see where the main argument is going and to give too much weight to the minor points made along the way.

A word of warning: reading philosophers' own words requires more time than you might expect. Unless you are truly exceptional, reading the words of John Locke, David Hume and Derek Parfit – the

The original title of the chapter from which this extract is taken was 'Of identity and diversity'. This appeared in the second and subsequent editions of Locke's *Essay* (the usual way of referring to the *Essay concerning Human Understanding*).

philosophers whose work features in the study associated with this book – will demand concentrated effort and a certain amount of re-reading. This is nothing like reading a newspaper, or indeed like most novel reading. However, don't be deterred by the apparent difficulty when you start. The teaching materials are designed to help you to make sense of what otherwise might seem impenetrable writing. As with riding a bicycle or learning to drive a car, reading philosophy does get much easier as you progress – though if you find it *too* easy, you're probably missing something.

## 2.2 Preparing to read Locke

English that was written several centuries ago presents special problems for most students. Sometimes a translation of an early text originally written in a foreign language can be easier to understand than something written in archaic English. This is because many translators of such works translate into present-day English and so interpret the work for our time as they do so. Some German speakers allegedly prefer to read English translations of Immanuel Kant's notoriously difficult *Critique of Pure Reason*, for example, rather than attempting it in the original German, because the translators interpret the text for them and put it in a modern idiom.

Locke, however, was writing in the seventeenth century, and obviously our language has changed significantly since then, as have the conventions of philosophical communication. To understand why he wrote as he did, we also need some awareness of the historical context, particularly in this case relating to expectations about resurrection and an afterlife.

Immanuel Kant (1724–1804) made significant contributions to metaphysics and ethics and was one of the towering figures of the Enlightenment.

### Locke's terminology

An important theme in Locke's work generally is that philosophical confusions often arise from imprecise uses of language. We each attach idiosyncratic ideas to key terms, and as a result may completely misunderstand one another; or else our ideas are so vaguely formed that it is not clear at all what we mean by them.

In his discussion of personal identity, Locke uses a number of words in precise – but not always obvious – ways. Once you've grasped how he is using these terms the writing becomes much easier to understand. The most important terms are 'man', 'person', 'substance' and 'soul'.

#### Man

For Locke, a **man** is simply a member of our species, *Homo sapiens*, a human being. So to say that someone is the same man as they were ten years ago is simply to point out that there is continuity of their living biological body. (Locke uses 'man' as some people use 'mankind', i.e. to refer to either a man or a woman, though to accuse him of sexism would be anachronistic.) In this sense, I am the same 'man' I was as a 5-year-old boy (which sounds a bit odd, but that is how he uses 'man') since there is bodily continuity (though clearly not complete material

identity) between what I was physically then and what I am physically now. When, after the war in Iraq, the deposed leader Saddam Hussein was discovered in hiding, there was initially a question of whether the individual who was captured was really Saddam Hussein (rather than someone who looked a bit like him). Obviously, psychologically he had been changed by his experiences, since being deposed as leader of Iraq, but the question in Locke's terminology was whether or not there was bodily continuity between the man Saddam Hussein and the one they had found, i.e. was he the same biological individual, albeit to some extent physically and psychologically different from what he had been. The answer, in this case, was determined by a DNA test.

## Person

We use the word 'person' interchangeably with 'man' or 'woman', but that is not how Locke used the term. For Locke you can be the same man as you were ten years ago without being the same person: what determines whether someone is the same person as they had been earlier is purely continuity of consciousness via memory. So, for example, if I stole some sweets as a child and then completely forgot this, then the child who stole the sweets would not be the same *person* in Locke's terms as the person I am now. I am the same *man* (i.e. the same individual member of the species *Homo sapiens*, though grown older) as the boy who stole the sweets, but not the same *person*. He uses the word 'self' sometimes too as a synonym of 'person'. For him, 'person' refers to what we mean by 'I'.

Locke makes the point elsewhere that 'person' is a **forensic** term, one that relates to legal and moral questions of praise or blame-worthiness. The question of whether or not someone is the same person they once were is, for him, intimately tied to questions of culpability. His interest in this is in terms of moral responsibility, not only in this world but also at the Day of Judgement in the next, when God allegedly sees into the minds of human beings and determines their fate according to their sins. For Locke, and his Christian contemporaries, the question about the Day of Judgement and what would happen subsequently was a genuine and pressing concern. This question drives most of his discussion of personal identity, both in terms of how God might judge individuals, and in terms of the question of what might reasonably be thought to survive bodily death.

You may well reject the idea of a Day of Judgement – you may, for example, be an atheist and find the idea that God exists mere



Figure 2.1 Stephan Lochner, *The Last Judgement*, c.1435, oil on wood, 125 x 174cm. Wallraf Richartz Museum, Cologne. Photo: © The Bridgeman Art Library.

superstition. Don't let that stop you from reading Locke closely. Much of what he writes about personal identity is independent of his views about God. Questions of personal identity have a wider significance than just being about what it is that, for example, a Christian believes will survive bodily death; they are questions about what we fundamentally are and where the limits of our moral responsibility lie.

## Substance

In ordinary language, a 'substance' is some kind of physical or chemical stuff. That is not at all how Locke uses the term. For him a substance can be immaterial (i.e. non-physical). Unless you realise this, reading Locke can be very confusing. He refers to the soul as an immaterial substance, which sounds like a contradiction in terms. What 'substance' means is anything that continues to exist without changing (and that does not depend on anything else to be what it is). A substance for Locke can be non-physical or physical: the term is neutral between the

two. So to say that a stone is the numerically same substance that it was ten minutes ago is to say that it is materially (i.e. physically) unchanged; but equally for Locke a soul or spirit (he uses the words interchangeably in the passage we will be looking at) can be the same substance that it was (even though it might sound strange to us to write of something non-physical being a substance).

## Soul

For Locke, a **soul** is an unchanging substance. It may or may not have a consciousness. When those who believe in reincarnation talk about the soul leaving one body and inhabiting another one, then they are using the term much as Locke used it. You might, on Locke's use of the term, have the same soul that you did forty years ago, even if you have complete amnesia and cannot recall anything at all that happened to you forty years ago.

---

## Activity

Turn to the extract from Chapter 27 of Locke's *Essay* in the set book (Part V, Reading 1, 'The Self and Consciousness', pp. 275–80). Read through the passage along with John Cottingham's brief introduction. The point of this activity is for you to get an overview and a general feel for Locke's writing. Don't worry if much of it makes very little sense to you at this point. Concentrate on the thought experiments, and any passages that do seem clear.

Once you have read quite quickly through to the end, read Jonathan Bennett's summary of the last part of this extract which is printed at the end of this book (Reading 1). This should help you get through some of the denser passages as it involves interpretation of the text. The aim here is to get a rough sense of this difficult passage. Once you have done that, we will be concentrating on just five important paragraphs towards the end of the extract that contain the kernel of Locke's thinking on the topic. Do not spend more than twenty minutes on this activity.

Have a look at the *Exploring philosophy* website for useful links to online material.

With out-of-copyright texts, such as the readings from John Locke and David Hume for this book, you may also be able to locate the set passages in online books, either by doing a search yourself or by following the suggestions in the 'Optional further material' pages on the *Exploring philosophy* website, which often includes links to exactly this kind of resource. This should give you more flexibility in your study.

---

## 2.3 Focusing in on Locke's *Essay*

We are now going to concentrate on five paragraphs of the Locke extract, as these contain his most important ideas about personal identity. Once you have studied these paragraphs in some depth, you may want to attempt to re-read the entire extract contained in the set book.

### Activity

Turn back to the extract 'The Self and Consciousness' in the set book (Part V, Reading 1). Re-read the five consecutive paragraphs from 'Self is that conscious thinking thing ...' (at the bottom of p. 278) to '... his conscience accusing or excusing him ...' (on p. 280). You may find it useful to number these five paragraphs in your copy of the book. (These are the paragraphs that are paraphrased in Reading 1 at the end of this book.)

### Activity

Now re-read the first of these five paragraphs (from 'Self is that conscious thinking thing ...', on p. 278, to '... as everyone who reflects will perceive', on p. 279). This should be your third reading of it. Then answer this question:

What is the thought experiment of the detached little finger meant to show?

### Discussion

Locke imagines the fanciful situation of someone's little finger being severed, and that person's consciousness remaining in the detached little finger. In such a situation, he believes, we would say that the little finger would be that person's self and the remaining body, presumably now devoid of sensation and consciousness, would not be part of that self. This is meant to show that what counts as the person is determined by self-consciousness and memory rather than by body.



Figure 2.2 Little Finger. Photographed by DAJ. Photo: © amana images inc./Alamy.

### Activity

Re-read the second paragraph (from 'In this personal identity is founded ...' to '... imputed to him', on p. 279) and answer the following question:

What does Locke have to say about the situation in which, once the little finger had separated from the rest of the body, the rest of the body had its own consciousness?

## Discussion

If the new consciousness in the body was independent of the consciousness in the severed little finger, then that new consciousness would be a separate person and not the concern of the little finger. The point of this slightly surreal thought experiment is to emphasise that what makes each of us a person is the psychological continuity provided by memory and self-consciousness (i.e. *not* bodily continuity, though they usually coincide) – the main point of Locke's whole chapter.

---

## Activity

Re-read the third paragraph (from 'This may show us ...' to '(for such twins have been seen)', on p. 279) and answer the following questions before going on to the discussion below.

- 1 What does Locke mean by the phrase 'identity of consciousness' here?
- 2 Under what conditions would Locke be prepared to say that Socrates and the present mayor of Quinborough were the same person?
- 3 What point is Locke making when he talks about 'Socrates waking and sleeping'?
- 4 In this paragraph, what evidence is there that Locke is primarily concerned with 'person' as a **forensic** term?

## Discussion

Locke manages to pack several ideas into this short paragraph.

- 1 'Identity of consciousness' for Locke means 'sameness of consciousness', where consciousness is the self-reflective aspect of our experience. His point is that sameness of, for example, soul substance, doesn't make us the same person over time. Locke thinks the idea that a non-conscious soul might be what makes us the same person over time is mistaken. It is self-awareness, given by our memories of what we have done, as well as by our current experience that is the source of personal identity. It is not given by continuing physical substance – bodily continuity – either (a point emphasised in the previous paragraphs with the thought experiment involving the little finger).
- 2 Socrates would be the same person as the current mayor of Quinborough if the mayor had the consciousness of being Socrates by having memories of doing what Socrates did. The fact that the mayor of Quinborough had a different body and lived many centuries after Socrates would be irrelevant – identity of consciousness is what counts for Locke.

- 3 The thought experiment in which Socrates waking and Socrates sleeping are two distinct consciousnesses is Locke's way of showing that sameness of body doesn't guarantee personal identity: for him there are two persons associated with this single Socrates body. He emphasises this point by going on to assert that it would in those circumstances be wrong to punish the Socrates waking for what Socrates sleeping thought. The implication is that we should punish the correct person, and we identify persons via consciousness rather than bodies, or – as Locke elsewhere points out – that is what God would do.
- 4 The discussion of the justification for punishing Socrates waking in this paragraph is a clear indication that Locke is concentrating on forensic questions here: he is interested in questions of moral (and possibly legal) responsibility for actions and thoughts rather than in purely metaphysical questions (i.e. about the nature of reality) of whether or not Socrates waking is the same person as Socrates sleeping.

---

### Activity

Re-read the fourth paragraph (from 'But yet possibly it will still be objected ...' to '... was no longer in that man ...', on p. 279) and answer these questions before moving on to read the discussion below.

- 1 What is the objection that Locke considers in this paragraph?
- 2 What is his response to this objection?

### Discussion

- 1 Locke anticipates the objection that most of us completely forget some parts of our lives, yet surely each of us remains the same person as the person who lived through those forgotten days. His point is that this is inconsistent with his view that a person is determined by what they are conscious of: if someone irretrievably forgets something they have done, it must follow from Locke's theory that they are no longer the same person as the one who performed that action.
- 2 Locke's reply to this anticipated objection is to embrace the conclusion that forgotten parts of our lives are not part of us as persons. He claims that this is commonly recognised when a mad man is not punished for what he did when he wasn't mad (presumably because he could not remember it, and so on Locke's account would not be morally responsible for it, because he was not the same person). Locke's interpretation of phrases such as 'he was

not himself when he did that' is that they recognise to some degree that you can be dealing with the same man (in Locke's sense) but not the same person.

---

### Activity

Re-read the fifth paragraph (from 'But is not a man drunk ...', on p. 279, to '... his conscience accusing or excusing him ...', on p. 280) and answer the following question.

Why does Locke think it acceptable for human beings to punish a sober man for the actions he performed when drunk?

### Discussion

In this paragraph, Locke says that when someone claims not to remember what they did when drunk, human laws justly punish the individual for the drunken behaviour. The reason why this punishment is appropriate, he says, is that human beings are not in a position to determine whether or not the accused person really has forgotten what he did. Human beings can't prove that their claimed amnesia is genuine. In contrast, Locke maintains that come the Day of Judgement ('the Great Day'), God would not punish people for what they did not remember doing. In other words, human judges don't give drunks the benefit of the doubt if they claim when sober to have forgotten what they did; but a divine judge would be able to determine the truth of the matter and would not punish someone for acts they had genuinely forgotten.

---

## 2.4 Engaging with Locke

So far we've only been concerned with understanding what Locke meant in this passage. By this stage you should have a very good idea of the main themes that Locke brought out here. But studying philosophy is not just a matter of understanding what other people have said in the past – an activity that usually involves weighing evidence in favour of a particular interpretation. Studying philosophy is also a matter of *engaging* with the ideas of past and present thinkers. So what we are going to do now is to think critically about whether or not Locke's account is plausible. Are there any counter-arguments against his main position?

### Thomas Reid's criticism of Locke

According to Locke, the core of our personal identity is the reach of consciousness extending back. In particular, personal identity is dependent on our memory of having performed certain actions in the past. He is explicit that when we cannot remember our earlier actions as our own, we cease to be responsible for them from a legal and moral point of view (at least when our souls are bared before God – actual courts may judge us culpable for actions that we say we cannot remember because they have no straightforward way of assessing whether or not we are telling the truth about our forgetfulness, and aren't prepared to give us the benefit of the doubt).

However, as the Scottish philosopher Thomas Reid pointed out in the eighteenth century, Locke's account of the role of memory here leads to some very strange conclusions. Reid made this point effectively by imagining a quite plausible situation of gradual memory loss:

Suppose a brave officer to have been flogged as a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and have been major general in advanced life; suppose, also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that, when made general, he was conscious of his taking the standard, but had absolutely lost consciousness of his flogging. These things being supposed, it follows from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it

Thomas Reid (1710–1796) is perhaps best known for his *An Inquiry into the Human Mind, on the Principles of Common Sense* (1764).

A.C. Grayling and Paul Snowdon discuss this issue in the audio recordings 'Personal identity' (which you have already listened to) and 'Locke on persons' (which you will listen to later in this chapter).



Figure 2.3 Portrait of Thomas Reid 1796, oil on canvas. Fyvie Castle, Aberdeenshire. Photo: National Trust for Scotland.

follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging; therefore, according to Mr Locke's doctrine, he is not the person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school.

(Reid, 2002 [1785], p. 276)

So the old general can remember his earlier bravery as a young man, but has completely forgotten about the incident in which he stole apples from the orchard and was flogged for it. Even if someone were to remind him about the flogging, if he still could not genuinely remember it as something that he had done, then, according to Locke, he was no longer the same *person* as the boy who stole the apples. Yet, also according to Locke, because he could still remember his bravery as a young man, he was still the same person as that young man. As a point of logic, if A is identical to B, and B is identical to C, then A must be

identical to C. But, as Reid points out, this is not what Locke would believe in the case of the old general. If A = the old general, B = the brave young soldier, and C = the young boy, Locke says that A = B and B = C but on Locke's account A does not = C. As Reid puts it, 'the general is, and at the same time is not, the same person with him who was flogged at school'. This is a direct contradiction. The move of showing a contradiction that follows from a position taken by someone is known as a *reductio ad absurdum* (this is the strict sense of the term, though it is sometimes used in a looser sense to mean any demonstration that absurd consequences follow from a stated position). If a position leads to a direct contradiction, we know that it can't be true.

There is another way of putting this. Reid is pointing out that identity is a *transitive* relation: if A = B and B = C, then A = C. (A relation is transitive when, if A bears that relation to B and B bears it to C, then A bears it to C. The relation 'is a sibling of', for example, is transitive: if A is a sibling of B and B is a sibling of C, then A is a sibling of C.) According to Reid, Locke's theory of personal identity is inconsistent with the transitivity of identity.

## Activity

How do you think Locke might have replied to Reid's objection outlined above?

## Discussion

It is clear from the passage from Locke that we have read that he would not allow that the old general was the same person as the child who stole the apples and was flogged for it in Reid's example. He was certainly the same man, because of bodily continuity as the same individual human being. And the young man was at the time he took the standard the same person as the child who stole the apples. But the old general's amnesia about his childhood means that he was not the same person as the child. This is like the sober man's amnesia about his actions while drunk (Locke, in set book, p. 279). Locke might have argued that personal identity is not the same as strict logical identity, so it is not a transitive relation (to use the relevant jargon). He would certainly have maintained that the old general was not the same person as the boy thief, as is apparent from his anticipation of a related objection in the fourth paragraph discussed above.

The trouble with the sort of answer that Locke would give is that cases of forgetting are everyday. We all forget much of what we have done, often irretrievably. Our memories overlap like the general's, the young brave soldier's and the boy's. A consequence of accepting Locke's view literally here would be that my life as a man (in his terms) is radically different from my life as a person. As a man I was once a baby, a young child, an adolescent, young adult and I am now middle aged. But my early childhood experience, most of which is lost to me now, is better thought of, according to Locke, as belonging to a single unit of consciousness that is separate from what I now think of as my self. This is counter-intuitive as an account of what a person is. Perhaps a better way of describing personal identity, though not one Locke ever adopted, would be to say that what makes me the same person as I once was is this pattern of overlapping memories and not simply that only what I can now remember having done counts as an aspect of my self. In the audio recording 'Locke on persons', which you will listen to later in this chapter, Paul Snowdon mentions this approach to personal identity as one that some present-day philosophers take.

## J.L. Mackie's reading of Locke

We will now consider a section from J.L. Mackie's book, *Problems from Locke* (1976). Mackie was probably best known for his work on ethics, but he also wrote extensively on Locke and on Hume.

---

### Activity

Turn to Reading 2, 'Mackie on the unity of consciousness', which you can find at the back of this book. Read it through once and then re-read the last half of the final paragraph (from the sentence that begins 'This brings in another factor, one's concern about my own possible future happiness or misery ...' to the end of the reading).

Try to answer the following question before moving on to the discussion:

How can Locke's theory of personal identity have implications for what we do in the future rather than just how we understand our relation to our past?

### Discussion

Mackie makes the point that we have a special interest in what happens to us. In his reading of Locke on personal identity, it is the future rather than the past that motivates the discussion. Locke is very much focused

on concerns about the self that will potentially be punished or rewarded by God on the Day of Judgement. Even though most of Locke's actual discussion is backward-looking, in that he is interested in the grounds for praise or blame of past actions and the role of memory, Mackie points out that knowing that my future self can be justifiably praised or blamed for what I do now can be a factor in determining how I shall behave. In other words, the concept of moral responsibility that animates Locke's discussion here extends to the self's future as well as back to the self's past. What binds past, present and future together is the continuity of consciousness. We have a special concern for the self that we will be in the future.

---

Whether or not we should have such an exaggerated concern about the happiness or misery of the person we will become is a theme to which we will return when we discuss Derek Parfit's radical ideas on this topic in Chapter 4.

## Activity

Now listen to the audio recording 'Locke on persons' in which Paul Snowdon, who is Professor of Philosophy at University College, London, discusses Locke's views. Parts of this interview are quite complicated. Be sure to listen to the entire recording at least twice before moving on. Snowdon's more detailed overview of Locke's approach should complement the overview that A.C. Grayling gave in the audio recording 'Personal identity' (which you listened to in Chapter 1) and help you consolidate what you have learned about Locke so far.

In particular, you should listen out for and make your own notes on:

- Snowdon's account of the context in which Locke was writing
- his gloss of key words Locke uses, especially 'person', 'man' and 'forensic'
- his discussion of Thomas Reid's objection and the problem of transitivity.

---

The audio recording 'Locke on persons' lasts about fifteen minutes – but you will need to listen to it more than once.

## 2.5 Reading historical texts

What then are we doing as philosophers when we read texts like the extract from Locke and those you will encounter throughout your study of philosophy?

So far in this book we have been reading work by the seventeenth-century philosopher John Locke. But you might, legitimately, ask, 'Why are we doing this?' If you were studying physics, you might talk about Isaac Newton's ideas, but you probably wouldn't read original texts by him, and your main motivation would be to see how his ideas have been developed and superseded by subsequent scientists. Historians, of course, would want to get back to primary texts, and read them in a deeply contextualised way, taking into account social, economic and political factors that influenced the way they were written, related writings by contemporaries, and so on.

There are different ways that philosophers approach writings of the past. Below are the basic positions.

- ***As history of ideas:*** Some philosophers are fascinated by the detail and context of past philosophical writing. They approach a reading of a historical text principally as historians of ideas, wanting to understand the texts in the terms that the author's contemporaries would have done. This approach provides a very good antidote to casual readings of texts that naively see philosophers of the past as addressing precisely the same problems that we are now. The historical approach also allows us access to the great works of the past that are part of our cultural history and the source and stimulus for much present-day philosophical discussion. To read a text in this way requires a detailed historical understanding of the writer's period and concerns.
- ***As part of a conversation:*** The philosopher René Descartes wrote in his *Discourse on Method* (1637):

I was aware that the reading of all good books is indeed like a conversation with the noblest men of past centuries who were the authors of them, nay a carefully studied conversation, in which they reveal to us none but the best of their thoughts.

(Descartes, 1911 [1637], p. 84)

Many present-day philosophers read the work of past philosophers (including the work of Descartes) in this spirit. From a strict historical point of view this is anachronistic, an inappropriately ahistorical approach. For a historian of ideas, the thought that a long-dead philosopher is part of an ongoing conversation as if brought back to life through the text may be dismissed as sentimentality. Yet, throughout the history of philosophy, this is how most philosophers have engaged with past philosophers. Alfred Whitehead described all philosophy as 'footnotes to Plato', by which he meant that Plato mapped out the key questions and how we might answer them, and all subsequent philosophers have just made minor amendments to his ideas. This is probably not an accurate picture of philosophy, but it does bring out how vibrant and alive long-dead philosophers can seem to be: they continue to have the power to stimulate thinkers of the present day, even if present-day thinkers don't fully appreciate the historical circumstances in which the original writers were working.

There is a danger attached to a too-scholarly approach to the subject – as if the books themselves are more important than the ideas. If you lose the sense of reading philosophy because you want to think and philosophise yourself, and immerse yourself too deeply in the minutiae of the history of who said what, when and to whom for its own sake, you risk losing touch with the lifeblood of the subject. Arthur Schopenhauer wrote about this danger in his polemical essay 'On thinking for oneself', distinguishing the thinker from the mere scholar:

A truth that has merely been learnt adheres to us only as an artificial limb, a false tooth, a wax nose does, or at most like transplanted skin; but a truth won by thinking for oneself is like a natural limb: it alone really belongs to us. This is what determines the difference between a thinker and a mere scholar.

(Schopenhauer, 1970 [1851], p. 91)

My reading of this passage (which might turn out to be anachronistic) is that Schopenhauer has correctly identified a danger of a merely scholarly approach to philosophy: the point of studying this subject is because we are interested in the philosophical questions, not just to learn what some philosophers of the past happened to have said about these questions. Reading philosophers of the past truly can be like a conversation as we grapple with the ideas, try to understand and engage critically with them and, in the process, hopefully, are stimulated to ideas of our own, or at least emerge with a clearer conception of what is important and why.

Alfred Whitehead  
(1861–1947), English  
mathematician and  
philosopher.

Arthur Schopenhauer  
(1788–1860) is best  
known for his book *The  
World as Will and Idea*  
(1818).

A moderate compromise between history of ideas and philosophy as conversation is probably the dominant approach in philosophy departments in universities today. One of the great benefits of the exchange of research through books, articles and now the internet is that there can be a useful division of labour within philosophy. Those who wish to engage with past philosophers as part of an ongoing conversation can avoid the crassest anachronisms because the work of the best historians of ideas is readily available to them.

---

### Activity

The audio recording 'Reading historical texts' lasts about twelve minutes – but you will need to listen to it more than once.

Now that you have had a chance to grapple with an extract from Locke's *Essay concerning Human Understanding*, and begun thinking about reading philosophy written several hundred years ago, listen to the audio recording 'Reading historical texts'. In it, Susan James and Miranda Fricker, both philosophers at Birkbeck College, London, talk about why philosophy as a subject, in contrast to, for example, the sciences, draws so much on its past. They also discuss how much historical context is needed in order to make sense of philosophy that is several hundred years old. In the course of the discussion both philosophers make useful comments about Locke on personal identity, and about the general approach taken by the philosopher we are going to focus on in Chapter 3 – David Hume.

---

# Summary

The key points covered in Chapter 2 are:

- Locke believed that personal identity was a matter of consciousness, not bodily continuity.
- ‘Person’ for Locke was a forensic term – linked with questions of moral responsibility and punishment.
- Locke uses a variety of thought experiments to illuminate his ideas about personal identity.
- Thomas Reid criticised Locke because on Locke’s view personal identity proved not to be transitive.
- Reading historical texts is an important aspect of philosophical study that requires selective appreciation of context.

In this chapter we have focused on the skills of reading an extract from a seventeenth-century philosopher’s work, listening and note-taking from audio discussions, and engaging critically with philosophical ideas.

In the final two chapters of this book we will be looking at David Hume’s idea that the self is a fiction, and considering some of the present-day philosopher Derek Parfit’s radical ideas about the self and the future.

---

## Activity

Now have a look at the end-of-chapter quiz on the *Exploring philosophy* website.

---

## References

Descartes, R. (1637) *Discourse on Method* in Haldane, E. and Ross, G. (trans.) (1911) *The Philosophical Works of Descartes*, vol. 1, Cambridge, Cambridge University Press.

Reid, T. (2002 [1785]) *Essays on the Intellectual Powers of Man* (ed. D.R. Brookes), University Park, Pennsylvania, Pennsylvania State University Press.

Schopenhauer, A. (1970 [1851]) ‘On thinking for oneself’ in Hollingdale, R.J. (trans.) (1970) *Essays and Aphorisms*, Harmondsworth, Penguin Classics.

# Chapter 3

## The self as a fiction



# Contents

Aims	57
Materials you will need	58
3.1 Introduction	59
3.2 Hume's search for the self	61
3.3 A closer reading of the Hume extract	65
Hume on impressions and ideas	66
Returning to the extract: paragraphs 3 and 4	70
Change, continuity and identity	72
Could you have an idea without a corresponding impression?	74
Summary	75
References	76



# Aims

By the end of this chapter, you should:

- have read and understood the extract by David Hume in the set book
- have appreciated something of the context in which Hume was writing.

## Materials you will need

In this chapter you will need to refer to the following readings in the set book:

- Part V, Reading 3, 'The Self as Bundle: David Hume, *A Treatise of Human Nature*' (focusing on the first six paragraphs).

You will need to listen to the following audio recordings (available on the *Exploring philosophy* website):

- Hume: an introduction
- Hume on the self.

You will also be directed to the website for an end-of-chapter quiz.

## 3.1 Introduction

The identity which we ascribe to the mind of man is only a fictitious one.

(Hume (1739–40), in set book, p. 289)

On John Locke's account, we only have personal identity in the sense of psychological continuity via consciousness. For Locke, the continuity over time is achieved by persisting memories, though, as Thomas Reid showed, this leads to some strange and apparently contradictory conclusions. This doesn't mean that all accounts of personal identity based on continuity of consciousness or memory are necessarily flawed – only, perhaps, Locke's extreme view which required that we remember our past actions for them genuinely to be *our* actions. Reid, for his part, believed in an unchanging self that is not reducible to anything else. Locke, as you will recall, rejected the idea that a person was an unchanging soul substance; he also rejected the idea that our personal identity arose simply from bodily continuity on the grounds that we can imagine existing in a different body from the one that we have.

The practical upshot of this for Locke would have been that he could imagine surviving death in a body that was different from the one he had in life, and that even if consumed by a cannibal there would not be the problem of re-gathering all the separate ingested parts to ensure continuing post-mortem existence – something that deeply worried many of his contemporaries, including his friend the scientist and philosopher Robert Boyle. This may seem rather an exotic thought to us now, but in Locke's time there really was concern over how the body might be reconstituted after death to ensure continuing existence. That's one reason why the punishment of hanging, drawing and quartering someone's body was considered so pernicious – it made it much harder to get all the parts together come the Day of Judgement. Most present-day believers in immortality, in contrast, are more inclined to believe in the existence of a soul or immaterial self that survives bodily death rather than the resurrection of the actual body.

In this chapter, we will look at the eighteenth-century philosopher David Hume's attack on the notion that introspection (i.e. reflection on our own experience) gives us knowledge of a soul or self. Controversially, he thought that the self was a fiction. Hume does not believe that it makes any sense at all to talk of an unchanging self that

Robert Boyle (1627–1691) was a natural philosopher (what we would now call a scientist) who made important discoveries in chemistry and physics.

unifies the person. When he reflects on his own experience, he can find no enduring self within him at all, and assumes that no one else can either (though, with characteristic irony, he does allow that other people may find something in themselves that eludes him when he reflects on his own experience).

## 3.2 Hume's search for the self

### Activity

Now listen to the audio recording 'Hume: an introduction', which includes thoughts from A.C. Grayling, Simon Blackburn, Peter Jones and Peter Millican. This should give you further background to help you appreciate the work of an important philosopher whose work will be discussed throughout *Exploring philosophy*.

The audio recording 'Hume: an introduction' lasts about six minutes – but you will need to listen to it more than once.

Hume, who is very much within the empiricist tradition of philosophy following on from Locke, believed that the contents of our minds consisted of what he called **ideas** and **impressions**. Impressions were our sensory experiences that 'impress' on the mind. In Hume's terminology, I'm now having an impression of a computer screen. When I shut my eyes and visualise the screen, I have an 'idea' of the screen. Hume believed that all simple ideas come from some corresponding sensory experience. We learn from the senses and gradually build up our knowledge of the world. We will look more closely at this idea in the section 'Hume on impressions and ideas'.

Where, then, does the *idea* that we have of our own self come from? Hume is puzzled by this question and rejects the traditional answers.

### Who was David Hume?

Hume was one of the most impressive thinkers of the eighteenth-century Scottish Enlightenment and probably the most important philosopher to have written in English: the contemporary philosopher Simon Blackburn describes him as 'the greatest British philosopher' (2008, p. 1). Renowned in his day as a historian and essayist, Hume wrote on a very wide range of topics including economics. His multi-volume *A History of England* was for many years the definitive book on the subject. But it is as a philosopher that he is best known today – this despite his first book, his *Treatise*, having, in his words, fallen '*dead-born from the press*' (Hume, 2001 [1776], p. 612).

Hume developed Locke's empiricist ideas, constructing a science of humanity, a mixture of what we would now call psychology and philosophy. He was also at least an agnostic, and very probably an

You can hear Simon Blackburn talking about Hume in the audio recordings for this chapter.

You will explore Hume's *Dialogues concerning Natural Religion* in Book 2, Chapter 2.

atheist. His posthumous *Dialogues concerning Natural Religion* was a brilliant attack on many traditional arguments for God's existence, particularly the Argument from Design.

As well as being a penetrating thinker, Hume was also a very skilful and witty writer. This is one of the reasons why so many of his readers admire his work.



Figure 3.1 Allan Ramsay, *David Hume*, 1766, oil on canvas, 76 x 64 cm. Scottish National Portrait Gallery, Edinburgh. Photo: © The Bridgeman Art Library.

The famous extract from Hume's *Treatise* that we are going to read contains a sceptical account of the thought that there might be an enduring self, followed by a more constructive account of where our idea of an enduring self (which he thinks is a fictitious one) comes from. The idea of an enduring self, he acknowledges, is more or less universal. Just about everyone thinks they have a self that persists through time. But where is the corresponding sensory experience? When we introspect, we never encounter 'the self'; we never have an 'impression' of it in Hume's terminology. He is then left with the difficulty of explaining why confidence in the idea that we all have an unchanging self within us is so commonplace. It must, he thinks, be some kind of construction of the mind, not something which exists independently of our patterns of thought.

In other words, Hume not only rejects the idea that there is a self or soul underlying all our experience, but also acknowledges the temptation of thinking of ourselves in these terms. To explain the situation, he

provides a psychological explanation of why we might be seduced into believing that there is some such unchanging self lurking behind everything that we do. He remains, though, adamant that no such self exists. The unchanging self (or soul) is a fiction. In his discussion, like Locke, he gives an important role to memory; but whereas Locke thought that our memory *constitutes* our personal identity, Hume believed that memory *bundles* our experiences together and gives us the illusion that they all belong to a continuing self. For this reason, his theory is often described as the **Bundle Theory of the Self**, even though he doesn't believe that the bundle is an enduring self, but rather the source of the illusion that we have a self.

### Activity

Read the whole of Part V, Reading 3, 'The Self as Bundle: David Hume, *A Treatise of Human Nature*' in the set book, including the introduction by Cottingham.

### Discussion

The point of this first reading is simply to get an overview of Hume's writing on this topic. Don't worry if you found this difficult. We will be focusing on a key passage from the reading.

### Activity

Re-read the first four paragraphs of this reading in the set book (from 'There are some philosophers who ...' on p. 285 to '... the materials of which it is composed' on p. 286). Remember that the title of this extract in the set book has been given by John Cottingham and is not Hume's own – this comes from Book 1, Part IV, Section 6 of Hume's *Treatise*.

At this stage you should be reading to get a sense of the main ideas, rather than trying to understand every sentence. As presented by Cottingham, Hume is confidently asserting that the self is a fiction. However, it is worth noting that in an appendix to the *Treatise* he indicates that he is unhappy with his account of how we achieve an idea of the self (perhaps the reason that he omitted it from his later reworking of the *Treatise*, the *Enquiries concerning the Human Understanding*). A further reason might have been that it could easily have been read as a direct attack on religion since orthodoxy would claim the existence of an enduring soul – nevertheless, the *Enquiries* included a critique of the Argument from Design and a sceptical essay on reports of miracles, so there is circumstantial evidence that this wasn't the reason for the omission).

## Discussion

This part of the reading contains a negative conclusion: the claim that we have no direct evidence of an enduring self when we reflect on our own experience. This is the central theme of the section of the reading we are focusing on here. In the rest of the extract, Hume hypothesises about the likely causes of a widespread belief in what he takes to be a fiction: the enduring self.

---

## Activity

The audio recording 'Hume on the self' lasts about ten minutes – but you will need to listen to it more than once.

Listen to the audio recording 'Hume on the self' which features Simon Blackburn, author of *How to Read Hume*. This should help you to understand the key features of Hume's approach and his conclusions about the self. Like the interview with A.C. Grayling (that you listened to during your study of Chapter 1) it also includes some discussion of Derek Parfit's view of the self (the topic of next week's work). Both audio recordings should be very useful in consolidating your learning when you have reached the end of this book and when you are preparing for the examination.

---

## 3.3 A closer reading of the Hume extract

### Activity

Now, re-read the first paragraph of Part V, Reading 3, 'The Self as Bundle: David Hume, *A Treatise of Human Nature*' in the set book (from 'There are some philosophers who ...' to '... if we doubt of this', on p. 285).

In this paragraph, Hume lines up his target: philosophers who believe that we are always aware of our self. Most of Hume's readers would have recognised immediately that he had in mind Joseph Butler, since not only had Butler vigorously asserted the existence of an unchanging self, but he also claimed that the notion of such a self was somehow primitive (note this is the title Cottingham gives Reading 2 in Part V of the set book: 'The Self as Primitive Concept: Joseph Butler, *Of Personal Identity*'), 'beyond the evidence of demonstration' as Hume puts it (set book, p. 285). Butler certainly believed that we are 'intimately conscious' (Hume's words, in set book, p. 285) of our continuity personal identity grounded in an unchanging self:

every person is conscious that he is now the same person or self he was as far back as his remembrance reaches, since when anyone reflects upon a past action of his own, he is just as certain of the person who did that action, namely himself – the person who now reflects on it – as he is certain that the action was *at all* done.

(Butler (1736), in set book, p. 284)

The way Hume sets this up in this first paragraph suggests that he is simply going to attack this view. In fact, as will emerge, having knocked this view down, because he believes it is contradicted by experience, he acknowledges that it is a very widespread feeling that such a self exists, and he goes on to diagnose why people might have come to think in this way.

Joseph Butler (1692–1752) was an English bishop, theologian and philosopher.

---

### Activity

If you want to find out more about Butler's view of the self, read the extract from *The Analogy of Religion* (1736) in the set book (Part V, Reading 2, 'The Self as Primitive Concept'). Otherwise, go on to the next activity.

---

### Activity

Now read the second paragraph of the reading by Hume in the set book (from 'Unluckily all the positive assertions ...', on p. 285, to '... consequently there is no such idea', on p. 286).

---

Hume maintains that the claim that we have an unchanging self is just contrary to experience. The terms he uses to make this point need some exposition as they relate directly to his general empiricist approach: his account of where the contents of the mind come from. Put simply, sensory experience is the source of our thoughts. All knowledge must be derived from the senses. It is worth digressing here to give the background to Hume's approach in this passage.

## Hume on impressions and ideas

For Hume, the word 'impression' is a technical term with a precise meaning. Impressions are our sense experiences. Look around you – you are getting sense impressions of your surroundings. Listen to the noises around you – I can hear music playing and people talking, though I can't make out what they are saying. The experiences I am having are all, in Hume's language, 'impressions'. He presumably chose that word because it is as if the world, via the senses, imprints itself on our mind – that it leaves an impression as a footprint in the snow does. Furthermore, he believes that impressions resemble the things they are representations of.

Now close your eyes and recall what you could see or hear. The thoughts that you are having are 'ideas' in Hume's terminology. If you see a cat, you have an impression of the cat; if you close your eyes and think about that same cat, you are having an idea of the cat. Ideas are

copies of impressions. All our ideas come from impressions, according to Hume. In his words:

all our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent.

(Hume, 1969 [1739–40], p. 52)

Subjectively, the impressions feel stronger – they are more vivid, or ‘lively’ as Hume puts it – and the corresponding ideas are faint copies of these impressions. Before I ever heard the sound of laughter, I could not have an idea of laughter. As Hume puts it in his *An Enquiry concerning Human Understanding*:

A blind man can form no notion of colours; a deaf man of sounds. Restore either of them that sense, in which he is deficient; by opening this new inlet for his sensations, you also open an inlet for the ideas; and he finds no difficulty in conceiving these objects.

(Hume, 2008 [1748], p. 14)

Here he is thinking of someone totally blind or totally deaf from birth. On this view I couldn’t have an idea of red if I had never seen some red thing at some point in my life; nor could I have an idea of the taste of wine if I had never tasted it. Having the experience of drinking it is a precondition of my having an idea of its flavour. I can, Hume thinks, form a complex idea of a gold mountain, even though I’ve never seen one, by combining simple ideas of ‘gold’ and of ‘mountain’. That is what we do when we imagine things that we have never seen. But I could never get the idea of the colour gold without perceiving some gold thing – that’s because the gold colour is, in his view, a simple idea and so requires a prior impression in order for me to have an idea of it.

---

## Activity

Read the following passages from Hume’s *Treatise* and *Enquiry*.

The first passage is from the opening of Hume’s *Treatise* – the section called ‘Of the origin of our ideas’. Here Hume explains how he uses the words ‘impression’ and ‘idea’. Impressions and ideas are both *perceptions* for Hume, i.e. they are both thoughts – the former are sense experiences (what we would be more likely to call perceptions) and the latter are fainter representations of our sense experiences and feelings.



Figure 3.2 The taste and smell of wine comes via our senses. Photographed by Jan Caudron. Photo: © Jan Caudron/Anaklasis/Alamy.

All the perceptions of the human mind resolve themselves into two distinct kinds, which I shall call IMPRESSIONS and IDEAS. The difference betwixt these consists in the degrees of force and liveliness, with which they strike upon the mind, and make their way into our thought or consciousness. Those perceptions, which enter with most force and violence, we may name *impressions*; and under this name I comprehend all our sensations, passions and emotions, as they make their first appearance in the soul. By *ideas* I mean the faint images of these in thinking and reasoning; such as, for instance, are all the perceptions excited by the present discourse, excepting only, those which arise from the sight and touch, and excepting the immediate pleasure or uneasiness it may occasion.

(Hume, 1969 [1739–40], p. 50)

The second passage is from Hume's *Enquiry*. This should help you make sense of what Hume believes he is doing in the extract about personal identity, namely revealing that the idea of a continuing self requires an impression that just isn't there. He is also diagnosing the source of the idea that so many people have of there being a continuing self.

All ideas, especially abstract ones, are naturally faint and obscure: the mind has but a slender hold of them: they are apt to be confounded with other resembling ideas; and when we have often employed any term, though without a distinct meaning, we are apt to imagine it has a determinate idea annexed to it. On the contrary, all impressions, that is, all sensations, either outward or inward, are strong and vivid: the limits between them are more exactly determined: nor is it easy to fall into any error or mistake with regard to them. When we entertain, therefore any suspicion, that a philosophical term is employed without any meaning or idea (as is but too frequent), we need but enquire, from what impression is that supposed idea derived?

(Hume, 2008 [1748], p. 15)

---

Much of Hume's discussion of personal identity and the idea that we allegedly have of an unchanging soul is devoted to asking just this question: from what impression is that supposed idea derived? Hume's resolute answer, backed by the findings of his introspection, and made in stylish prose, is that we have no impression whatsoever of a soul or other unchanging self-substance. The concluding lines quoted above rely on a notion traditionally known as Hume's Fork. This is the view that if an idea is neither derived from an impression, nor true by definition, then it is worthless.

## Hume's Fork

In one of the most famous passages in his *Enquiry concerning Human Understanding* Hume outlines his approach to philosophy. Take any bit of argument or reasoning, Hume says, and there are just two options: it must either be true by definition, as is  $2 + 3 = 5$ , or else empirically discoverable (i.e. its truth or falsity can be discovered by some kind of observation). If it is neither, then, he says, it is 'sophistry and illusion' and should be burned. Here is the passage:

If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number? No. Does it*

*contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.*

(Hume, 2008 [1748], p. 120)

Now we shall return to the opening of the second paragraph of Part V, Reading 3, 'The Self as Bundle: David Hume, *A Treatise of Human Nature*' in the set book, which starts 'Unluckily all these positive assertions ...'. Having declared that we don't have 'any idea of *self*', Hume asks the question 'For from what impression could this idea be derived?' (set book, p. 285). This is Hume's way of saying that, assuming any idea of self is a *simple* idea (i.e. not a *complex* one deriving from multiple impressions), if it genuinely occurred there would have to have been some sensory impression which brought it about and of which it would be a faint copy. In other words, Hume is completely committed to his empiricist thesis that all our ideas come from impressions, so he asks from which impression (sensory experience) any idea of self might arise. This is because he has, he believed, proved elsewhere in his *Treatise* the thought that he expresses here: 'It must be some one impression that gives rise to every real idea' (set book, p. 285). If there were a simple sensory input giving rise to the idea of self, then it would have to be a constant, running without change throughout our lives. But, as Hume points out, that is not the case. No one has a 'constant and invariable' impression: rather we experience a succession of impressions, including pain, pleasure, and grief and joy. These things come one after another, and none of them remains steady providing a source for an idea of the unchanging self. But for Hume, if there is no simple impression that brings about the idea, then 'there is no such idea' (set book, p. 286).

## Returning to the extract: paragraphs 3 and 4

### Activity

Read the third and fourth paragraphs of the reading by Hume in the set book (from 'But further, what must become ...' to '... the materials of which it is composed', on p. 286). This is one of the most famous passages in Hume's writing. Here he is describing the fruits of his introspection. He cannot discover any self that lurks behind his experience. All that he finds are fleeting experiences, one after another.

In this passage, Hume begins to ask difficult questions, questions which take him towards the surprising view that the self is an illusion. We all have a succession of perceptions, but what joins them together making them *my* perceptions and not someone else's or even no one's at all?

The passage beginning 'For my part ...' is Hume's eloquent description of what he finds when he reflects on his own experience – namely, always a perception of some kind or other, a feeling, a sight or smell of something, a thought. He can never catch himself with no thought at all, nor can he get a glimpse of the elusive unchanging self that thinkers such as Butler declare they can perceive. When he sleeps soundly, Hume points out, he has no thoughts at all, and so in a sense ceases to exist.

At death, if all his thoughts and feelings cease, Hume believes, then he too would cease to exist. There are those who claim that a person's soul might survive death and the destruction of all consciousness (this is one view of the soul that is reincarnated that some Hindus subscribe to), but Hume can't understand them. For him it is clear that loss of perception would be annihilation. Hume does, though, with some irony, allow that some people may be able to perceive their unchanging souls within them, even though he cannot. To understand why he adds this as a possibility, recall that in his day such views would have been considered sacrilegious – there was very widespread belief in the notion of a soul. Hume was suspected of being an atheist (a plausible hypothesis) or at least an agnostic. Certainly, his belief that the self was an illusion was, from a religious point of view, highly unorthodox.

What then joins together the perceptions that we have, making them part of one person's consciousness? What provides the link between the various thoughts and feelings that Hume has observed? What makes them all *his* perceptions and not someone else's? Hume describes the 'theatre' of the mind, with a quick succession of ideas and impressions passing across it (though he is careful not to take this metaphor too far, since there is no 'stage' in the mind, only the succession of perceptions). There is no obvious substance bundling the various perceptions together. Why then do we think of ourselves as individuals with a continuing existence over time when all that seems to be there is a series of thoughts, feelings, impressions, and so on? For one very important answer to this question, listen again to Simon Blackburn talking about Immanuel Kant in the audio recording 'Hume on the self'.

## Change, continuity and identity

---

### Activity

Read the fifth and sixth paragraphs from the reading by Hume in the set book (from 'What then gives us so great a propensity ...', on p. 286, to '... our notion of identity', on p. 287).

---

Hume's answer to the question of where we get our idea of personal identity draws on a distinction between strict identity and the kind of identity over time that we attribute to animals and plants. We've already encountered a similar opposition of types of identity in reading Locke, so these notions should be familiar to you by now.

Paragraph six opens with a description of 'identity' or 'sameness' in a strict sense. Something that remains the same over time is identical with itself. This is the least problematical sense of identity. As well as appreciating strict identity of this kind, we also have a notion of objects that consist of material that changes over time, but for which there is close continuity over that change. This is really a succession of related objects, but we often think of such objects as identical over time, even though strictly speaking they are not.

Here Hume gives his diagnosis. The experience of perceiving an unchanging object over time and that of perceiving an object that, through a succession of linked stages, changes somewhat, can be very similar. It can be as if we are perceiving an unchanging object in both cases: think of a tree changing over time. We look at it and it seems to be the same tree that it was yesterday, though in fact much about it has changed. Leaves have dropped off or been devoured by caterpillars, the bark has peeled slightly more, the roots have grown a few millimetres, and so on. The experience of looking at it is similar to that of perceiving something that does not change over time. Almost unconsciously we begin to think of the tree as strictly identical with itself over time, when we know at some level that it is not. We have a natural bias to see objects that change gradually through a succession of stages as if they were identical with their earlier instantiations, despite our knowing that they are not. To justify this to ourselves, we often invent explanations as to why a changing thing is really identical with what it was before, even though at some level this is an absurd claim.

Hume suggests that the notion of soul, or some unchanging self-substance, is just such a useful device to disguise our mistake of describing things that are different and merely succeed each other as if they were all part of one identical thing.

David Hume's suggestion that the enduring self is a fiction is a challenging one. Just because it is a view that a famous philosopher held, it doesn't mean that that view is correct. It may well result from flawed reasoning or be based on a false assumption. If the reasoning is faulty, there is no guarantee that the conclusion is true; similarly if the argument is based on a false assumption, the conclusion may or may not be true. Coming to see how any philosophical view can be questioned, in terms of both its reasoning and its author's assumptions, is an important part of understanding that view. As you progress through this course you will frequently be asked to reflect critically on the material you are studying. The following activity questions an assumption that Hume made.

---

## Activity

Try thinking this through yourself. Ask yourself 'Do I ever have a direct experience of my self?' Sit down for a few minutes and reflect on the experiences you are having. Is it really true that, as Hume believed, you have no impression of your self, but just a succession of fleeting experiences? Or is there something behind all those experiences?

## Discussion

Here you are being asked to investigate one of Hume's assumptions, namely that his experience of the quest for a self generalises to everyone. You might find yourself agreeing entirely with Hume that all you can find is fleeting experiences and no underlying self. Some people, however, claim to be able to find one when they introspect like this. If Hume is right about the self, then they must be wrong. There may be some impression that they've tracked down, but it can't be the impression of the self that they've found. Are you convinced by alleged sightings of the impression of the self? Or do you think that Hume is right to say that there are no such sightings?

---

## Could you have an idea without a corresponding impression?

Hume, as we have seen, was completely committed to the notion that every simple idea derives from an impression. This is the core of his **empiricism**. In other words, he did not believe that we could have any innate ideas: we learn from experience. But what if he was wrong about this? If innate ideas are a possibility, then, perhaps, the source of the widespread belief in the existence of an enduring self is simply an inborn idea, not derived from any experience. Perhaps all human beings are born with this innate sense of self. That is one possibility that Hume did not seriously consider as a solution to the question 'Where does the idea of self come from?' That's because accepting it would have entailed his abandoning his fundamental beliefs about the origin of ideas.

Another possible response to Hume's position is that the self is real, but just isn't the kind of thing that there *could* be an impression of. Remember Hume's metaphor of the mind as a theatre. He might be right that the self is not a thing that ever appears on the stage in that mental theatre. But that doesn't necessarily mean that it isn't real. Maybe the self is like the audience. Or maybe it is like the theatre itself – a structural pre-condition of anything entering the stage at all. That was roughly Kant's position.

Contemporary philosophy is still very concerned with the important question of how our knowledge relates to our experience, and this is a theme that will resurface in later parts of *Exploring philosophy*, too.

# Summary

In this chapter we have examined David Hume's famous claim that when he introspected he could not find an enduring self, only a succession of impressions and ideas.

This week's work involved the skills of reading and engaging with an extract from an eighteenth-century text.

In the next chapter of this book, we'll be looking at the contemporary philosopher Derek Parfit's ideas about the self. Parfit combines the sort of enthusiasm for using thought experiments that we saw in John Locke's work with David Hume's scepticism about the idea of an enduring self, to produce a highly distinctive account of the self and the future.

---

## Activity

Now have a go at the end-of-chapter quiz on the *Exploring philosophy* website.

---

## References

Blackburn, S. (2008) *How to Read Hume*, London, Granta Books.

Hume, D. (1969 [1739–40]) *A Treatise of Human Nature* (ed. E.C. Mossner), Harmondsworth, Penguin Classics.

Hume (2001 [1776]) 'My Own Life', reprinted as an appendix to Mossner, E.C. *The Life of David Hume*, 2nd edition, Oxford, Clarendon Press.

Hume, D. (2008 [1748]) *An Enquiry concerning Human Understanding*, Oxford, Oxford University Press, World's Classics.

# Chapter 4

## The self and the future



# Contents

Aims	81
Materials you will need	82
4.1 Introduction	83
4.2 About my future	85
The duplicate that survives death	86
4.3 'Beam us up Scotty!'	88
How relevant are thought experiments?	92
Reading Parfit in the set book	94
4.4 Parfit's view and the significance of death	95
4.5 Criticising Parfit: Charles Taylor on the unity of a life	97
Summary	99
References	101



# Aims

By the end of this chapter, you should:

- understand Derek Parfit's general approach to questions about persons
- have thought critically about the use of thought experiments in this area.

## Materials you will need

In this chapter you will need to refer to the following readings in the set book:

- Part V, Reading 5, 'Liberation from the Self: Derek Parfit, *Reasons and Persons*'
- Part V, Reading 6, 'Selfhood and Narrative Understanding: Charles Taylor, *Sources of the Self*'.

You will need the following additional reading, which can be found at the end of this book:

- Reading 3: Vesey interviews Parfit on personal identity.

You will need to listen to the following audio recordings (available on the *Exploring philosophy* website):

- Parfit: an overview
- Parfit on personal identity.

You will also be directed to the website for an end-of-chapter quiz and other online activities.

## 4.1 Introduction

The truth is very different from what we are inclined to believe.

(Parfit (1984), in set book, p. 298)

At this point in the book we will leap forward more than two hundred years from David Hume's eighteenth-century scepticism about the existence of an enduring self to Derek Parfit's much more recent analysis of the self and the future. Parfit, born in 1942, is an Oxford-based philosopher who has been influenced by both John Locke and by Hume. He has brought a highly original angle to questions about persons. His thoughts in this area were first made public in a series of articles published in the 1970s. These led to the book *Reasons and Persons* (1984) which many philosophers see as one of the most important works of that decade (as I write, his second book is in press). Whether or not they agree with his conclusions, most philosophers recognise that Parfit is one of our most innovative thinkers. Like Locke, he makes extensive use of thought experiments, some of them quite elaborate.

---

### Activity

Listen to the audio recording 'Parfit: an overview'. This is an interview with Helen Beebee who is Professor of Philosophy at Birmingham University. It is intended to give you an overview of the main features of Parfit's approach to questions about the self. When reading Parfit's work it can be quite difficult to appreciate what his overall stance is. Listening to this interview should help you understand what is distinctive about this approach. I recommend that you listen to it at least three times. This may sound excessive, but Beebee compresses quite a few ideas into the space of just over nine minutes.

Write answers to the following question before reading the discussion below.

Helen Beebee summarises Parfit's approach in terms of two main aspects, a positive and a negative one. What are these aspects?

The audio recording 'Parfit: an overview' lasts about ten minutes – but you will need to listen to it several times.

### Discussion

The first point Beebee discusses – the positive aspect – is Parfit's view that questions about survival are more interesting and informative than questions of personal identity. Does anything of the 8-year-old Helen

Beebee survive in the present-day Helen Beebee? She maintains that there is very little of the 8-year-old personality now in her. In contrast, there is a great deal of continuity between how she was on the day before the interview and how she was on the day of the interview. The two kinds of relationship are very different. Suppose we can answer the question 'Am I the same person I was when an 8-year-old?' Beebee suggests the answer is likely to be 'Yes', but that very little informative hinges on that. Far more interesting and relevant are questions about the psychological and other continuities that hold between the earlier and later self. Personal identity questions aren't very important questions to ask, even if we can give answers to them.

The second aspect Beebee discusses – the negative one – is Parfit's point that in many imaginary cases there is no right answer to the question 'Is this person now the same person as person x at an earlier time?'

---

## 4.2 About my future

### Activity

Listen to the audio recording ‘Parfit on personal identity’, which is an archival recording of Derek Parfit (or rather Parfit’s former self – since the interview was first broadcast in February 1973) discussing personal identity with Godfrey Vesey (a former Professor of Philosophy at The Open University). You may find it useful to read the transcript which is reprinted at the back of this book (Reading 3) as you listen, or to read the transcript separately.

The audio recording ‘Parfit on personal identity’ lasts about twenty minutes – but you will need to listen to it more than once.

Parfit clearly believes that some kind of psychological continuity is what matters in questions about personal identity. This, you will remember, is, in general terms Locke’s view, too. Parfit, however, uses the thought experiment of brain bisection to illustrate the philosophical complexities that arise when there are two candidates for being the same person as the one that existed before the brain operation. In this thought experiment – which, though science fictional now, isn’t so far from becoming a possibility at some stage – my brain is split and put into two different people’s brainless skulls (this is the case that Helen Beebee describes in the audio recording ‘Parfit: an overview’ which you listened to earlier). The patients are then sewn up again, and re-awakened. Each of them remembers the whole of my life, has the same character as me, and is psychologically continuous with me in every way. If I’m the one going in for the operation, in such a case, Parfit points out, there are three answers we can give to the question ‘What’s going to happen to me?’

The three answers are these:

- (a) I’m going to be both the resulting people.
- (b) I’m going to be just one of the resulting people.
- (c) I’m going to be neither of the resulting people.

### Activity

Now answer these two questions:

- 1 What does Parfit believe to be wrong with each of these three answers?
- 2 What conclusion does he draw?

## Discussion

- 1 (a) Parfit rejects the idea that I will be both of the resulting people on the logical grounds that these are two distinct people and a person can't be two people leading different lives. Think how strange it would be to say that I am identical to (i.e. the same person as) someone who has a degree and with someone who doesn't have one. This is not how we use the word identity.
- (b) He dismisses the second answer, that I will be just one of the resulting people, because it would be 'wildly implausible' to make the choice between the two as I have the same relation to each. There aren't, then, good grounds for preferring one above the other. As Helen Beebee says in the audio recording 'Parfit: an overview', it would be completely arbitrary to make a choice between the two.
- (c) The third option, that I am going to be neither of the resulting people, Parfit opposes as 'grossly misleading', in that it implies that having such an operation would be as bad as dying. But, he maintains, it clearly isn't that bad. Think of what we would say if just one person survived the transplant – it would seem appropriate to say that I survived the operation. So how can having two people survive the operation (the case imagined) be worse than that?

- 2 The conclusion that Parfit draws from this is that we can imagine situations in which the question 'Will that future person be me or someone else?', first, doesn't have an answer', and, second, does not leave a puzzle to be solved. Where more than one candidate for being identical with me exists, as in this thought experiment, we can't, he believes, talk about personal identity. That's because identity requires a one-one relationship not a one–two relationship. This seems very strange at first because we are so used to assuming that 'Will I continue to exist?' is a meaningful question in every imaginable case. Here, though, we have an imagined case of full psychological continuity, but no personal identity. What we're left with is, as Parfit points out, *survival* without identity. But nevertheless the survival seems important.

---

## The duplicate that survives death

A second thought experiment Parfit uses is this: I die naturally, but scientists construct a perfect physical duplicate of me out of organic matter. This duplicate wakes up and has all my memories, and is psychologically continuous with me in every way.

---

## Activity

What does Parfit believe this thought experiment reveals? How does he reinforce this point?

## Discussion

Parfit's point is that when we ask the question 'Will the new-bodied person be me or someone else?', the two answers we can give ('yes' and 'no') are just two ways of describing the same events.

This is hard to take in, and Parfit realises this. We are used to thinking in one or other of two ways, neither of which sits comfortably with Parfit's suggestion. Obviously, as Godfrey Vesey points out in the audio recording 'Parfit on personal identity', some people think that personal identity is a matter of having the same immaterial soul. Then the question would be answered by discovering somehow whether or not the soul was transferred to the new body – though it is by no means clear how anyone would ever check that. Notice that Parfit doesn't want to countenance such a view. For those, in contrast, who believe that personal identity is a matter of being a physically continuous animal the answer would be that the animal's death begins a process of decay that results in it ceasing to exist. On that view, no replica, whatever its psychological properties, could be the same person.

In order to make his view more plausible, Parfit describes another case, one in which he thinks we are inclined to think correctly (as he sees it) about identity. The case is of a club that ceases to meet and then is resurrected by two of its former members fifteen years later in another country. Were we then to ask the question 'Have they started the same club or is it a completely new one?', Parfit suggests this would be an absurd question since it wouldn't make any difference at all. Analogously, he argues, the true view about personal identity is in the duplicate-body case he describes.

---

## 4.3 ‘Beam us up Scotty!’

The extract from Derek Parfit included in the set book is taken from his book *Reasons and Persons* (1984). We are only going to look at part of it (and you will read this later on in the chapter) – the sections ‘Why our identity is not what matters’ and ‘Liberation from the self’. In this form, it summarises some of his main conclusions, but does not always show the reasoning that leads to the position he takes. For a full understanding of why he believes what he does you will need to read this extract in the context of the whole book. But don’t go and do that now. For our purposes, a brief summary of some of his main views should suffice. Parfit’s writing is complex and in places extremely difficult to follow. It is also controversial. The main aim here is to give a sense of his approach and some of his most important conclusions.

Part Three of *Reasons and Persons* is devoted to the topic of personal identity. Other parts are more focused on morality and its relation to time. Part Three includes some very useful pointers to Parfit’s beliefs. In fact, the chapter headings of Chapters 10–13 are very useful for orienting yourself before reading the passage in the set book. Here are the chapter headings:

- Chapter 10: What we believe ourselves to be
- Chapter 11: How we are not what we believe
- Chapter 12: Why our identity is not what matters
- Chapter 13: What does matter

It should be clear from this that, like many philosophers (including, famously, Plato, who thought that the world that we inhabit and experience is a pale imitation of reality), Parfit believes appearances about what we fundamentally are to be deceptive. We believe ourselves to be one thing (heading to his Chapter 10) but in fact we are not that sort of thing at all (Chapter 11 heading). He goes on to say that, despite our tendency to worry about whether we will be the same person in the future, identity doesn’t really matter (Chapter 12 heading). What’s more, realising this can change how we think about such matters as ageing and death (Chapter 13 heading).

What is the view he rejects? Well, one aspect of it is this. Most of us most of the time believe that there is a definite answer to the question ‘Am I about to die?’, when there is some question about whether the

being that is about to die is me. Parfit thinks there need not be a determinate answer to this question. There can be cases, both imagined and real, where there is no 'right' answer to this question. How can this be? The easiest way to understand this is to think about a thought experiment he uses, based on an idea in the cult science-fiction television series *Star Trek*.



Figure 4.1 The transporter system in operation from *Star Trek*, 1966–69.  
Photo: Paramount Television/The Kobal Collection.

In *Star Trek*, crew members of the starship *Enterprise* can travel swiftly to the surface of any planet they are orbiting by means of a teletransporter. This apparently causes them to de-materialise from the spaceship and re-materialise on the planet surface (from which Scotty, the chief engineer, can beam them up again to the mothership when they have finished their work). Parfit uses a thought experiment based on this idea of a teletransporter to develop his ideas about persons. Here is Parfit's variant:

I enter the Teletransporter. I have been to Mars before, but only by the old method, a space-ship journey taking several weeks. This machine will send me at the speed of light. I merely have to press

the green button. Like others, I am nervous. Will it work? I remind myself what I have been told to expect. When I press the button, I shall lose consciousness, and then wake up at what seems a moment later. In fact I shall have been unconscious for about an hour. The Scanner here on Earth will destroy my brain and body, while recording the exact states of all of my cells. It will then transmit this information by radio. Travelling at the speed of light, the message will take three minutes to reach the Replicator on Mars. This will then create, out of new matter, a brain and body exactly like mine. It will be in this body that I shall wake up.

Though I believe that this is what will happen, I still hesitate. But then I remember seeing my wife grin when, at breakfast today, I revealed my nervousness. As she reminded me, she has been often teletransported, and there is nothing wrong with her. I press the button. As predicted, I lose and seem at once to regain consciousness, but in a different cubicle. Examining my new body, I find no change at all. Even the cut on my upper lip, from this morning's shave, is still there.

(Parfit, 1984, p. 199)

### Activity

Imagine you are about to enter Parfit's teletransporter described above and 'travel' to Mars. Write down brief answers to these questions before reading the discussion below.

- 1 Will you die?
- 2 Is the person who emerges from the machine on Mars really you?  
Yes? No? Not sure?
- 3 If you answered 'Yes' to question 2, explain why. If you answered 'No', explain why not.

### Discussion

Parfit's teletransporter thought experiment is in the tradition of personal identity thought experiments that stems from Locke. It brings out clearly the difference between those who think that personal identity is a matter of having bodily continuity (this is the view that Paul Snowdon, who appears in the audio recording 'Locke on persons' that you listened to during Chapter 2, takes); and those who believe that it is the character, or memories, that principally determine whether or not a person continues to exist, irrespective of the body.

If you answered that entering the teletransporter would result in your death, that entering willingly was a kind of suicide, then you probably believe that what makes you you is a matter of having this particular body rather than one indistinguishable from it. Bodily continuity without a break is key for this view. You can remember getting up this morning, no doubt, and your body now is pretty much continuous with the body that got out of bed – there are some cells missing, some new ones, too, but the physical history of how your body now got to where you are from where you were could, in principle, be very easily traced. There won't be any gaps; no times when your body didn't exist between this morning and now; your body won't have spontaneously disappeared in one place and reappeared mysteriously somewhere else. Locke would presumably say that if you entered the teletransporter the body that arrived on Mars would not be the same *man* as the one that entered the machine, since his notion of a man (a human being) requires bodily continuity. Not so for him the notion of a person.

But you might have answered that the individual who emerged on Mars would be you. If you take this view, you can't consistently believe that personal identity requires bodily continuity in any ordinary sense. More likely, you, like Locke, and many other philosophers, hold that what makes you you is the pattern of overlapping memories that you have.

You may not entertain precisely Locke's view, since Thomas Reid's objection to it is telling (see Chapter 2); but some psychological rather than merely bodily features must be the basis of personal identity on this view. The individual on Mars has all your memories (and what's more, looks exactly like you did, cell for cell, when you entered the machine); on that basis you are implying that what matters in personal identity is psychological rather than bodily continuity. In daily life your bodily and psychological continuity go together – most people don't experience teletransportation or body swaps of the prince/cobbler variety (very probably nobody does). But the thought experiment brings out what is at stake.

---

Parfit thinks such questions may not have right answers. He also thinks that the fact that there might not be right answers about personal identity in these sorts of cases (and in less exotic ones) shouldn't worry us unduly. He takes a position that has many parallels with Buddhism. Simply put, the idea of a continuing self is an illusion – what we have are successive selves, some of which may only be very distantly related to what we are now, so distantly related that it might be irrational to care more about your future self than about some other person altogether.

## How relevant are thought experiments?

W.V.O Quine (1908–2000) was an influential American philosopher, famous for his work in logic and the philosophy of language.

As mentioned in the audio recording on the topic ('Thought experiments', listened to in Chapter 1), some philosophers are sceptical about the benefits of using this sort of thought experiment in philosophy. Parfit relies very heavily on thought experiments in his arguments. He anticipates the objection and quotes W.V.O. Quine, an eminent twentieth-century philosopher, who took that line and questioned the value of appealing to intuitions in such unusual circumstances:

The method of science-fiction has its uses in philosophy, but ... I wonder whether the limits of the method are properly heeded. To seek what is 'logically required' for sameness of person under unprecedented circumstances is to suggest that words have some logical force beyond what our past needs have invested them with.

(Quine quoted in Parfit, 1984, p. 200)

Quine was by no means alone in having reservations about the use of science-fiction thought experiments in the area of personal identity. The Oxford philosopher Kathleen Wilkes (1946–2003) was also, for example, highly critical of this sort of device. In particular she argued that we have empirical evidence of a range of interesting real-life cases including split-brains and dementia that are far more informative about the nature of the self than these imagined and often highly implausible scenarios (Wilkes, 1988).

This sort of objection would also apply to most of John Locke's thought experiments discussed earlier. Asking questions about such imaginary cases may not be very useful if the circumstances are so different from anything we have previously met with and are never likely to encounter in real life. It is pushing language too far to take our reactions to such science-fiction cases as having any import.

---

### Activity

Listen again to the interview with Julian Baggini, in the audio recording 'Thought experiments', before turning to these questions.

- 1 List three of Locke's thought experiments.
- 2 Do you think Quine's criticism of thought experiments holds for these three?
- 3 How do you think Parfit might answer Quine's criticism?

## Discussion

- 1 Locke's writing on personal identity includes numerous thought experiments, including the prince and the cobbler, the day- and the night-man, and so on. Look back at Chapter 1 of this book to check that your three thought experiments come from Locke.
- 2 All Locke's thought experiments are certainly far-fetched in that they are unlikely to occur in reality (although there does seem to be some cases of genuine multiple personalities inhabiting a single body). If Quine is right that such thought experiments push the use of words beyond what is reasonable, then this does undermine Locke's use of them. It is difficult, though, to know what Locke could use in place of thought experiments to investigate the sorts of questions that interested him.
- 3 Parfit considered and responded to Quine's objection in *Reasons and Persons*. The passage is worth quoting:

The criticism might be justified if, when considering such imagined cases, we had no reactions. But these cases arouse in most of us strong beliefs. And these are beliefs, not about our words, but about ourselves. By considering these cases, we discover what we believe to be involved in our own continued existence, or what it is that makes us now and ourselves next year the same people. We discover our beliefs about the nature of personal identity over time. Though our beliefs are revealed most clearly when we consider imaginary cases, these beliefs also cover actual cases, and our own lives.

(Parfit, 1984, p. 200)

Parfit's point is that such thought experiments are means to discover what we actually believe about personal identity over time. If we had no strong beliefs when presented with such far-fetched cases, then, perhaps, Quine's criticism would be fair. But Parfit takes the fact that we have strong beliefs about such cases as evidence that they are performing this job of revealing our intuitions about personal identity to us.

You don't have to agree with Parfit here. Some philosophers agree with Quine and Wilkes that examining science-fiction examples is not the best way to proceed when trying to unravel questions of personal identity and that writers like Parfit put too much weight on such examples and what we might say about them.

## Reading Parfit in the set book

---

### Activity

Now read the two sections from the extract from Derek Parfit in the set book (Part V, Reading 5, ‘Liberation from the Self: Derek Parfit, *Reasons and Persons*’), entitled ‘Why our identity is not what matters’ (on p. 297) and ‘Liberation from the self’ (pp. 298–9). Also read Cottingham’s introduction to the reading (pp. 296–7). You may need to read sentences, paragraphs or indeed the whole extract several times before it makes any sense at all. Bear in mind as you read that Parfit subscribes to what he labels ‘the reductionist view’ and opposes what he calls ‘the non-reductionist view’ (see **reductionism about personal identity** in the glossary).

### Discussion

For Parfit, ‘the non-reductionist view’ is the default position: it is probably the view that most of us have prior to engaging in philosophical speculation about these sorts of issues (and possibly after this, too). Parfit not only rejects this view, but suggests that this rejection may lead to a radical revision of how we think about our own future, and in particular our own death. He doesn’t pretend that believing his conclusions will be easy to do: in fact he recognises that they are deeply counter-intuitive.

---

## 4.4 Parfit's view and the significance of death

---

### Activity

Return to the extract from Parfit in the set book (Part V, Reading 5) and re-read the third and fourth paragraphs of the section 'Liberation from the Self' (from 'When I believed ...' to '... seems to me less bad' on p. 298). Write a short summary of the key ideas as a paragraph entitled 'Why Parfit cares less about his death than he used to'.

### Discussion

Here is my attempt to summarise these paragraphs. This is, again, quite a difficult passage. You may strongly disagree with what Parfit is saying here, too. The point is to try and understand why he might care less about his death now he has embraced his reductionist view about personal identity.

#### Why Parfit cares less about his death than he used to

Parfit used to hold the view that most of us hold, that we in some sense remain the same person throughout our lives. That made death seem quite bad as he believed that his current self would at some future date simply cease to exist. Now he doesn't believe this in quite the same way and is less worried by death. There will be a psychological and physical continuity between the experiencer that will not be around in the future (and will cease to exist), and some of Parfit's thoughts and intentions now will affect how things turn out. But this is all there is to the self: relations (which can be stronger or weaker in various respects), rather than something that continues to exist over time. Nor do all these relations necessarily cease at the point of death. Some relations can continue posthumously: other people will have memories of Parfit, and may be influenced by things he thought or wrote for a while. So from his perspective his own death is not as bad as it seemed before. It is not his present self that will cease to exist, but rather some experiences in the future will no longer be related in quite the same way to his present experiences. He finds consolation in thinking about the future in this way.

For someone who, unlike Parfit, holds the view that what makes us the same person over time is that we are the same animal, that we are primarily physical beings above all, this won't provide much consolation. Death of an animal is at least the start of ceasing to exist (the corpse may continue to exist for a while, but will eventually decompose). I have a special conscious and unconscious attachment to *this* body and its future as it gradually deteriorates over time and eventually changes to dust. In particular, I want it to continue living, albeit in a gradually changing form (provided it isn't in extreme pain) as long as feasible.

---

## 4.5 Criticising Parfit: Charles Taylor on the unity of a life

Not everyone is convinced by Parfit's approach to questions of personal identity. We have already seen that some philosophers have doubts about his extensive use of science-fiction thought experiments and the conclusions he draws on the basis of them. The contemporary Canadian philosopher Charles Taylor makes a similar point in passing, but the main thrust of his criticism of Parfit derives from his view that it is an unavoidable aspect of being human that we orient ourselves in relation to what he calls 'the good' – the values that shape what we are and what we strive to become.

Charles Taylor (b. 1931) is a Canadian philosopher, well known for his book *Sources of the Self* (1989).

### Activity

Read the extract from Taylor's *Sources of the Self: The Making of Modern Identity* (1989) in the set book (Part V, Reading 6, 'Selfhood and Narrative Understanding') and Cottingham's introduction to the extract (on pp. 302–3). Write a short answer to this question:

What is Taylor's main criticism of Parfit's approach?

Even if you don't fully understand Taylor's criticism, you should still attempt to answer this question before reading the discussion that follows.

### Discussion

Parfit's approach explains the self in terms of psychological connectedness and/or continuity with the right cause. But Taylor rejects this 'neutral' or 'bleached' self as not capturing the role of human concerns, the things we value, in shaping a self. In Taylor's example, to me, my car is a whole single thing, but to a mechanic, it is a combination of discrete parts – when you ask what is the car 'really' in itself, there is no right answer here. In contrast, for Taylor, a self is not like a car. It is not arbitrary whether we see it as a single thing or a set of connected functional parts. For Taylor, a human is not a neutral object but is in part constituted by the concerns he or she has, and these concerns ultimately link to the values we measure ourselves against. His conclusion is that we have to understand ourselves in terms of whole lives. He goes on to think about this in relation to narrative, the story of our lives, in which a later good act can redeem an earlier failing. This is all part of making sense of our lives. As he puts it 'to repudiate my childhood as unredeemable in this sense is to accept a kind of mutilation as a person'

(Taylor (1989), in set book, p. 306). This view is in stark contrast to the one that Helen Beebee put forward about her relationship to her childhood in the audio recording 'Parfit: an overview', where she was giving a sympathetic account of Parfit's stance on our present relationship to our pasts.

---

Taylor believes that there is something close to an 'a priori unity of a human life through its whole extent' (set book, p. 307). 'A priori unity' here means that it is not just a matter of our experience happening to be of the different phases of our lives connecting up to form a narrative, but rather that the unity is a structural pre-given: it has to be that way. He holds back from saying that it is an actual 'a priori unity' because (see the second paragraph on p. 307) he can imagine a culture in which there was a fundamental split of individuals' lives into two parts. That imaginary case aside, though, a consequence of Taylor's view is that we have to see our lives in relation to concerns that shape a life as a whole.

---

### Activity

Now re-read the whole extract from Taylor. The process of trying to summarise his main criticism of Parfit, and then reading my summary of it should have made it far easier to grasp what Taylor is saying when you re-read it.

---

# Summary

In this chapter, you have focused on getting an overview of some of Derek Parfit's ideas about the self. The key skill here is that of making sense of quite difficult prose and of picking out the most important thoughts. You have also examined some possible criticisms of Parfit's approach.

Throughout the book, we have been looking at contrasting answers to the question 'What am I?' John Locke's writing on personal identity has set the framework for much subsequent debate. He made the case for a certain sort of psychological continuity being the core of personal identity. This contrasts strongly with the idea that a person is simply a biological organism: the idea that physical rather than psychological continuity is what produces our identity over time.

Locke used the words 'man' and 'person' in a precise way to clarify his view that it would, in principle, be possible for an entity to be the same man, but not the same person (i.e. the same animal, but not the same individual from a psychological point of view). He also contested the idea that what makes us the same person is that we have a continuing soul-substance.

As well as making the case for psychological continuity as the source of personal identity, Locke made fruitful use of a wide range of thought experiments, such as the imagined cases of the day- and the night-man, and the prince and the cobbler. These carefully controlled imaginary cases were not decorations to his work, but the main way in which he argued and clarified his position: they were designed to show that we can tease apart these notions of a human being as an animal and a human being as a person. Furthermore, he stressed the link between our notion of a person and our beliefs about moral responsibility: this is what he meant when he labelled 'person' a 'forensic' term. We hold people responsible for what they have done: controversially he thought that we should, ideally, only hold people morally responsible for what they could remember doing (though he conceded that from a practical angle it would not be appropriate to take someone's word that they had forgotten a crime as conclusive evidence that they had). Thomas Reid's objections to Locke's memory criterion of identity – that someone might remember their early childhood when middle-aged, and their middle age when elderly, but might forget their childhood when elderly, yet still remain in an important sense the same person as the child – are

important. Modern philosophers who accept some psychological criterion of personal identity usually modify their account to take heed of Reid's objection.

Locke's views contrast sharply with those of Joseph Butler. He was certain that he could perceive a continuing soul within himself. David Hume, after thinking hard about the issue, and observing his thoughts, could find no such self. He diagnosed the source of the idea of a self in terms of the way ideas get joined together. For him the enduring unchanging self was a fiction: instead there is just a bundle of fleeting thoughts.

The contemporary philosopher Derek Parfit draws on the traditions of seventeenth- and eighteenth-century theorising about the nature of personal identity in coming up with his surprising conclusions about the self. He believes that we are all to some degree in thrall to the idea that we have an unchanging self, but that this is an illusion. To that extent he shares Hume's view. But, in focusing on questions about our feelings about the future, he goes beyond Hume. Looking forward to my future, I shouldn't be unduly concerned about what will happen to the self that will be descended from the one that I am now – there will be continuities and differences, and there is a real sense in which that future self won't be the same as me, despite my tendency to think that it will. In contrast, Charles Taylor argues that Parfit neglects the close relationship between our sense of our selves and questions about the meaning and purpose of our lives as a whole.

---

### Activity

Listen again to the audio recording 'Personal identity', the interview with A.C. Grayling which you listened to in the first chapter. Now that you have studied all the thinkers that he mentions, his overview should make much more sense to you and help you to consolidate your work on this topic before you move on to Book 2. It should also be helpful later for revision before the examination.

As a final activity, you should take the end-of-chapter quiz on the *Exploring philosophy* website.

---

## References

Parfit, D. (1984) *Reasons and Persons*, Oxford, Clarendon.

Wilkes, K.V. (1988) *Real People: Personal Identity Without Thought Experiments*, Oxford, Clarendon.



# Readings



## Contents

Reading 1	Locke on personal identity (Bennett's paraphrase)	107
Reading 2	Mackie on the unity of consciousness	110
Reading 3	Parfit on personal identity	115



# Reading 1 Locke on personal identity (Bennett's paraphrase)

Source: Bennet, J. (2007), paraphrase of Locke, J. (1690) *An Essay concerning Human Understanding*, Bk II, ch. 27, paragraphs 17–22; available at [www.earlymoderntexts.com](http://www.earlymoderntexts.com) (Accessed 18 June 2010).

Some formatting has been removed.

17 Self is that conscious thinking thing that feels or is conscious of pleasure and pain and capable of happiness or misery, and so is concerned for itself as far as that consciousness extends. (This holds true whatever substance the thinking thing is made up of; it doesn't matter whether it is spiritual or material, simple or compounded.) You must find that while your little finger is brought under your consciousness it is as much a part of yourself as is your head or your heart. If the finger were amputated and this consciousness went along with it, deserting the rest of the body, it is evident that the little finger would then be the person, the same person; and this self would then have nothing to do with the rest of the body. As with spatial separation so also with temporal: something with which the consciousness of this present thinking thing can join itself makes the same person, and is one self with it, as everyone who reflects will perceive.

18 Personal identity is the basis for all the right and justice of reward and punishment. What everyone is concerned for, for himself, is happiness and misery – with no concern for what becomes of any substance that isn't connected with that consciousness. [Locke goes on to apply that to his 'finger' example, supposing that the finger takes the original consciousness with it, and that the rest of the body acquires a new consciousness.]

19 This illustrates my thesis that personal identity consists not in the identity of substance but in the identity of consciousness. If Socrates and the present mayor of Queenborough agree in *that*, they are the same person; if Socrates awake doesn't partake of the same consciousness as Socrates sleeping, they aren't the same person. And to punish Socrates awake for something done by sleeping Socrates without Socrates awake ever being conscious of it would be as unjust as to punish someone for an action of his twin brother's merely because their outsides were so alike that they couldn't be distinguished.

20 It may be objected: ‘Suppose I wholly lose the memory of some parts of my life beyond any possibility of retrieving them, so that I shall never be conscious of them again; aren’t I still the same person who did those actions, had those thoughts that I once was conscious of, even though I have now forgotten them?’ To this I answer that we must be careful about what the word ‘I’ is applied to. This objector is thinking of sameness of the *man*, and calls it ‘I’ because he assumes that the same man is the same *person*. But the assumption isn’t necessarily correct. If one man could have distinct disconnected consciousnesses at different times, that *same man* would certainly make different *persons* at different times. That this is what people in general think can be seen in the most solemn declaration of their opinions: human laws don’t punish the madman for the sane man’s actions, or the sane man for what the madman did, because they treat them as two persons. This is reflected in common speech when we say that someone is ‘not himself’ or is ‘beside himself’. Those phrases insinuate that the speaker thinks – or that those who coined the phrases thought – that the self was changed, the self-same person was no longer in that man.

Note that this paragraph has been omitted from the reading in the set book.

21 ‘It is still hard to conceive that Socrates, the same individual man, might be two persons.’ To help us with this we must consider what is meant by ‘Socrates’, or ‘the same individual man’. There are three options. The same man might be any of these:

- 1 the same individual, immaterial, thinking substance; in short, the numerically-same soul and nothing else.
- 2 the same animal, without any regard to an immaterial soul.
- 3 the same immaterial spirit united to the same animal.

Help yourself! On any of these accounts of ‘same man’, it is impossible for personal identity to consist in anything but consciousness, or reach any further than that does.

According to 1, a man born of different women, and in distant times, might still be the same man. Anyone who allows *this* must also allow that the same man could be two distinct persons. ...

According to 2 and 3, Socrates in this life cannot be the same man as anyone in the after-life. The only way to do this – allowing for the possibility that Socrates in Athens and Socrates in Limbo are the same man – is through an appeal to sameness of consciousness; and that amounts to equating human identity – ‘same man’ – with personal identity. But that equation is problematic, because it makes it hard to

see how the infant Socrates can be the same man as Socrates after the resurrection. There seems to be little agreement about what makes a man, and thus about what makes the same individual man; but whatever we think about that, if we are not to fall into great absurdities we must agree that sameness of person resides in consciousness.

22 You may want to object: 'But isn't a man drunk and sober the same person? Why else is he punished for what he does when drunk, even if he is never afterwards conscious of it? He is just as much a single person as a man who walks in his sleep and is answerable, while awake, for any harm he did in his sleep.' Here is my reply to that. Human laws punish both, with a justice suitable to the state of knowledge of those who administer the law: in these cases they can't distinguish for sure what is real from what is counterfeit; and so they don't allow the ignorance in drunkenness or sleep as a plea. Granted: punishment is tied to personhood, which is tied to consciousness, and the drunkard may not be conscious of what he did; but the courts justly punish him, because his bad actions are proved *against* him, and his lack of consciousness of them can't be proved *for* him. It may be reasonable to think that on the great day when the secrets of all hearts are laid open, nobody will be held accountable for actions of which he knows nothing; everybody will receive his sentence with his conscience agreeing with God's judgment by accusing or excusing him.

## Reading 2 Mackie on the unity of consciousness

Source: Mackie, J.L. (1976) *Problems from Locke*, Oxford, Oxford University Press, pp. 173–7.

### Locke and the Unity of Consciousness

Identity in general is a problem for logicians and philosophers, but the identity of persons is of much wider concern. Personal immortality, survival after bodily death, and the transmigration of souls have at all times been objects of religious belief, anxious doubt, and intensely interested speculation. Cases of changed or divided personality and loss of memory are prominent in imaginative literature as well as in popular psychology, and science fiction has introduced further fascinating or horrifying possibilities. Some of the actual or possible odd cases are not merely of private interest, but do, or might, call for legal decisions about penalties, responsibilities, duties, and rights. The philosophical problem of what constitutes the identity of a person is highly relevant to all of these. On the answer to it depends the very coherence of some of our hopes and fears as well as our understanding of and response to some of the things that do, or might, occur. It is not surprising, then, that it was to this problem that Locke devoted most of his chapter on identity and diversity, or that his answer, and the question itself, have continued to be keenly debated ever since.

Locke, in accordance with what we have called his principle of the relativity of identity, says that ‘to find wherein *personal identity* consists, we must consider what *person* stands for,’ and answers that this is ‘a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing in different times and places.’<sup>1</sup> But Locke is rightly concerned not just with the word ‘person’ itself. He is using it as the noun that corresponds to all the personal pronouns. His question is, ‘Wherein consists my identity, and hers, and his, and yours?’

His answer is that it is consciousness that constitutes personal identity, that makes me, for example, the same me, the same person, through and despite the passage of time. His argument is that consciousness ‘is inseparable from thinking’, that when we perceive or meditate or will we know that we do so, and that it is by this consciousness that each of

us considers himself as himself, as one persisting thinking thing. It is by this reflective consciousness that our different sensations and perceptions and thoughts and desires at any one time belong to one self, and, Locke thinks, the same principle must account for the sameness of the self at different times: 'as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that *person*: it is the same self now it was then, and it is by the same *self* with this present one that now reflects on it, that that action was done.'<sup>2</sup>

What this amounts to as a positive doctrine is not yet clear, but there are at least two views that Locke is plainly rejecting. First, in distinguishing the person from the man, and hence *the same person from the same man*, he is denying that bodily continuity, the persistence of the human organism, makes personal identity. The same living human body with its continuity of animal life constitutes the same man; but not necessarily, Locke holds, the same person. But secondly he is denying that to be the same person is to be, or to have, one persisting immaterial, spiritual soul-substance. Against this view he argues not by denying that there are spiritual substances but by saying that their identity does not matter. If there are soul-substances, then presumably these can be reincarnated: the present mayor of Queenborough may, for all that anyone knows, have what used to be the soul of Socrates; but if he has no consciousness from the inside of any of Socrates's actions or thoughts, no direct awareness of those experiences as his experiences, then he is not the same person as Socrates.<sup>3</sup>

Again, if the same soul-substance carried two alternating sets of co-conscious thoughts, there would be two different persons with one soul. Locke uses parallel arguments to bring out the irrelevance to the identity of the person of both the living body and the supposed soul-substance. 'Could we suppose two distinct incommunicable consciousnesses acting the same body, the one constantly by day, the other by night ... I ask ... whether the *day*- and the *night-man* would not be two as distinct persons as *Socrates* and *Plato*'. Similarly, since we know that an 'immortal thinking thing may sometimes part with its past consciousness and be restored to it again' – that is, there may be a total but temporary loss of memory – we can imagine in the same soul-substance 'these intervals of memory and forgetfulness to take their turns regularly by day and night', and then 'you have two persons with the same immaterial spirit, as much as in the former instance two persons with the same body'.<sup>4</sup>

Conversely, if we had ‘the same consciousness, acting by intervals, two distinct bodies’, Locke argues that this would be the same person in the two bodies, just as you may have the same man in two different suits of clothes; and this still holds if there are also two distinct immaterial substances. Personal identity might be ‘continued in a succession of several substances’; it might be ‘preserved in the change of immaterial substances ... as animal identity is preserved in the change of material substances’.<sup>5</sup>

Locke is using the analogy of the way in which the same vegetable or animal life is continued despite the metabolic processes that constantly replace the material components of an organism to argue that there could be a sort of spiritual metabolism, the same consciousness being passed from one soul-substance to another.

Equally, it might be passed from one body to another: ‘should the soul of a prince, carrying with it the consciousness of the prince’s past life, enter and inform the body of a cobbler as soon as deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince’s actions’. This example is used to drive a wedge between *the same man* and *the same person*; no one, Locke thinks, would say that this is the same man as the former prince.<sup>6</sup>

In these examples Locke introduces a method which has been taken over by many later contributors to this discussion, the construction of puzzle cases. The continuity of animal life in one body and a unified mental history or unity of consciousness normally go together, and the presence of a single immaterial soul-substances was presumed by Locke’s contemporaries normally to accompany these. But Locke has imagined a series of cases in which the identity and diversity of these three items, body, consciousness, and soul, do not all go together, but are separated and combined in various ways. In examples of the Jekyll-and-Hyde type the same body and the same soul are associated with two separate unities of consciousness, two mental histories not linked by mutual awareness. In Socrates and the mayor of Queenborough we may have the same soul but different bodies and disjoint consciousnesses. In the prince and the cobbler we have the same soul and the same consciousness successively in two different bodies. And so on. Locke claims that in all these various combinations it is plausible to say that we have the same person where and only where we have the same consciousness; the sameness of the living body is neither necessary nor sufficient to constitute the same person, nor is the sameness of a spiritual substance.

Locke makes out a strong case for both his negative theses, that personal identity is to be equated neither with the identity of a soul-substance nor with that of a man, that is, of a living human animal body. There is also considerable plausibility in at least the broad outline of his positive thesis, that personal identity is somehow to be equated with, or based on, the unity of consciousness. True, he could hardly claim, and he does not in fact claim, that everyone uses the phrase 'the same person' in accordance with his principles. He admits that 'in the ordinary way of speaking, the same person and the same man stand for one and the same thing'.<sup>7</sup>

Rather what he is saying is that if we find three distinct meanings for the terms 'spirit', 'man', and 'person', we shall in consequence let the identity of persons be determined in the way that he suggests. He reinforces this claim by saying that 'person' is 'a forensic term, appropriating actions and their merit'. That is, the sameness of a person is intended to carry with it legal and moral responsibility for actions; that is why it 'belongs only to intelligent agents, capable of a law, and happiness and misery'. This personality 'imputes to *itself* past actions, just upon the same ground and for the same reason that it does the present. All which is founded in a concern for happiness, the unavoidable concomitant of consciousness: that which is conscious of pleasure and pain desiring that that self that is conscious should be happy'.<sup>8</sup>

This brings in another factor, one's concern for one's own future happiness. I have a peculiarly intimate, egoistic concern about my own possible future happiness or misery, quite different from any altruistic or sympathetic interest that I may have in the well-being of others: hardly anyone literally loves his neighbour as himself. Locke assumes that this forward-looking concern fits in with the reflective self-ascription of past actions: the future self for whose well-being I now have this special concern is the one which will impute to itself whatever I now do and which will remember my present experiences from the inside. If this is so, then it is peculiarly appropriate, especially on deterrent grounds, but also on some other views about punishment, to tie responsibility to such a two-directional unity of consciousness. If I know that the future self for which I have this special concern will be punished for my wrong actions (which it will remember and impute to itself), this gives me a reason for now refraining from wrong actions. But if I do not refrain, and that future self is punished for them, it will remember them and associate the punishment with them and so,

through its like concern for a further future self, will be deterred from repeating those wrong actions. Backward-looking memory and action-ascription and forward-looking concern seem thus to go together, and in conjunction they supply the rationale for the use of the concept of an identical person as the bearer of responsibilities and rights. This is, I believe, the thought behind what Locke says here, though we have to read between the lines to find it.

## Notes

- 1 II. xxvii. 9.
- 2 II. xxvii. 9.
- 3 II. xxvii. 14. I am assuming that 'one who was persuaded his had been the soul of *Socrates*' in this section is the mayor of Queenborough referred to as possibly identical with Socrates in Section 19. The description 'in the post he filled, which was no inconsiderable one, he passed for a very rational man' might well apply to a mayor, though 'the press has shown that he wanted not parts or learning' is contrary evidence, since there seems to be no record of publications by a mayor of Queenborough in the relevant period. Cf. query by R. H. in *Locke Newsletter*, No. 4 (1973), pp. 43–4.
- 4 II. xxvii. 23; cf. II. i. 11–12.
- 5 II. xxvii. 23, 10, 12.
- 6 II. xxvii. 15.
- 7 II. xxvii. 15.
- 8 II. xxvii. 26; cf. II. i. 11.

## Reading 3 Parfit on personal identity

Source: Parfit, D. and Vesey G. (1974) 'Brain Transplants and Personal Identity', in Vesey, G. (ed.) *Philosophy in the Open*, Milton Keynes, The Open University Press, pp. 55–64.

### Brain Transplants and Personal Identity

Derek Parfit and Godfrey Vesey

#### Brain Transplants

In 1973 in the *Sunday Times* there was a report of how a team from the Metropolitan Hospital in Cleveland under Dr. R.J. White had successfully transplanted a monkey's head on to another monkey's body.<sup>1</sup> Dr. White was reported as having said, 'Technically a human head transplant is possible,' and as hoping that 'it may be possible eventually to transplant parts of the brain or other organs inside the head.'

The possibility of brain transplants gives rise to a fascinating philosophical problem. Imagine the following situation:

Two men, a Mr Brown and a Mr Robinson, had been operated on for brain tumours and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. Let us call the latter 'Brownson'. Upon regaining consciousness Brownson exhibits great shock and surprise at the appearance of his body. Then, upon seeing Brown's body, he exclaims incredulously, 'That's me lying there!' Pointing to himself he says, 'This isn't my body; the one over there is!' When asked his name he automatically replies 'Brown.' He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life, always describing them as events in his own life. Of Robinson's past life he evinces no knowledge at all. Over a period of time he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on, that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson.<sup>2</sup>

The next step is to suppose that Brown's brain is not simply transplanted whole into someone else's brainless head, but is divided in two and half put into each of *two* other people's brainless heads. The same memory having been coded in many parts of the cortex, they *both* then say they are Brown, are able to describe events in Brown's life as if they are events in their own lives, etc. What should we say now?

The implications of this case for what we should say about personal identity are considered by Derek Parfit in a paper entitled, 'Personal Identity.' Parfit's own view is expressed in terms of a relationship he calls 'psychological continuity.' He analyses this relationship partly in terms of what he calls '*q*-memory' ('*q*' stands for 'quasi'). He sketches the definition of '*q*-memory' as follows:

I am *q*-remembering an experience if (1) I have a belief about a past experience which seems in itself like a memory belief, (2) someone did have such an experience, and (3) my belief is dependent upon this experience in the same way (whatever that is) in which a memory of an experience is dependent upon it.<sup>3</sup>

The significance of this definition of *q*-memory is that two people can, in theory, *q*-remember doing what only one person did. So two people can, in theory, be psychologically continuous with one person.

Parfit's thesis is that there is nothing more to personal identity than this 'psychological continuity.' This is *not* to say that whenever there is a sufficient degree of psychological continuity there is personal identity, for psychological continuity could be a one–two, or 'branching,' relationship, and we are able to speak of 'identity' only when there is a one–one relationship. It is to say that a common belief – in the special nature of personal identity – is mistaken.

In the discussion that follows I began by asking Parfit what he thinks of this common belief.

The following discussion is a transcript of the audio recording 'Parfit on personal identity', which is available from the *Exploring philosophy* website.

## Personal Identity

VESEY Derek, can we begin with the belief that you claim most of us have about personal identity? It's this: whatever happens between now and some future time either I shall still exist or I shan't – and any future experience will either be my experience or it won't. In other words: personal identity is an all or nothing matter – either I survive or I don't. Now what do you want to say about that?

PARFIT It seems to me just false. I think the true view is that we can easily describe and imagine large numbers of cases in which the question 'Will that future person be me or someone else?' is both a question which doesn't have any answer at all, and that there's no puzzle that there's no answer.

VESEY Will you describe one such case?

PARFIT One of them is the case discussed in the correspondence material, the case of division, in which we suppose that each half of my brain is to be transplanted into a new body and the two resulting people will both seem to remember the whole of my life, have my character and be psychologically continuous with me in every way. Now in this case of division there were only three possible answers to the question 'what's going to happen to *me*?' And all three of them seem to me open to very serious objections. So the conclusion to be drawn from the case is that the question of what's going to happen to me, just doesn't have an answer. I think the case also shows that that's not mysterious at all.

VESEY Right, let's deal with these three possibilities in turn.

PARFIT Well, the first is that I'm going to be both of the resulting people. What's wrong with that answer is that it leads very quickly to a contradiction.

VESEY Yes, why?

PARFIT The two resulting people are going to be different from each other. They're going to live completely different lives. They're going to be as different as any two people are. But if they're different people from each other, it can't be the case that I'm going to be both of them. Because if I'm both of them, then one of the resulting people is going to *be* the same person as the other, to wit me.

VESEY Yes. They can't be different people and be the same person, namely me.

PARFIT Exactly. So the first answer leads to a contradiction.

VESEY Yes. And the second?

PARFIT Well, the second possible answer is that I'm not going to be both of them but just one of them. This doesn't lead to a contradiction, it's just wildly implausible. It's implausible because my relation to each of the resulting people is exactly similar.

Here, 'correspondence material' refers to the original study materials accompanying this interview.

VESEY Yes, so there's no reason to say that I'm one rather than the other?

PARFIT It just seems absurd to suppose that, when you've got exactly the same relation, one of them is identity and the other is nothing at all.

VESEY It does seem absurd, but there are philosophers who would say that sort of thing. Let's go on to the third.

PARFIT Well, the only remaining answer, if I'm not going to be both of them or only one of them, the only remaining possible answer is that I'm going to be neither of them. What's wrong with this answer is that it's grossly misleading.

VESEY Why?

PARFIT If I'm going to be neither of them, then there's not going to be anyone in the world after the operation who's going to be me. And that implies, given the way we now think, that the operation is as bad as death. Because if there's no one who's going to be me, then I cease to exist. But it's obvious on reflection that the operation isn't as bad as death. It isn't bad in any way at all. That this is obvious can be shown by supposing that when they do the operation only one of the transplants succeeds and only one of the resulting people ever comes to consciousness again.

VESEY Then I think we would say that that person was me. I mean we'd have no reason to say that he wasn't.

PARFIT On reflection I'm sure we would all think that I would survive as that one person.

VESEY Yes.

PARFIT Yes. Well, if we now go back to the case where both operations succeed ...

VESEY Where there's a double success ...

PARFIT It's clearly absurd to suppose that double success is a failure.

VESEY Yes.

PARFIT So the conclusion that I would draw from this case is firstly, that to the question 'What's going to happen to me?', there's no true answer.

VESEY Yes.

PARFIT Secondly, that if we decide to say one of the three possible answers, what we say is going to obscure the true nature of the case.

VESEY Yes.

PARFIT And, thirdly, the case isn't in any way puzzling. And the reason for that is this. My relation to each of the resulting people is the relation of full psychological continuity. When I'm psychologically continuous with only one person, we call it identity. But if I'm psychologically continuous with two future people, we can't call it identity. It's not puzzling, because we know exactly what's going to happen.

VESEY Yes, could I see if I've got this straight? Where there is psychological continuity in a one-one case, this is the sort of case which we'd ordinarily talk of in terms of a person having survived the operation, or something like that.

PARFIT Yes.

VESEY Now what about when there is what you call psychological continuity – that's to say, where the people seem to remember having been me and so on – in a one-two case? Is this survival or not?

PARFIT Well, I think it's just as good a survival, but the block we have to get over is that we can't say that anyone in the world after the operation is going to be me.

VESEY No.

PARFIT Well, we can say it but it's very implausible. And we're inclined to think that if there's not going to be anyone who is me tomorrow, then I don't survive. What we need to realize is that my relation to each of those two people is just as good as survival. Nothing is missing at all in my relation to both of them, as compared with my relation to myself tomorrow.

VESEY Yes.

PARFIT So here we've got survival without identity. And that only seems puzzling if we think that identity is a further fact over and above psychological continuity.

VESEY It is very hard not to think of identity being a further fact, isn't it?

PARFIT Yes, I think it is. I think that the only way to get rid of our temptation to believe this is to consider many more cases than this one case of division. Perhaps I should give you another one. Suppose that the following is going to happen to me. When I die in a normal way, scientists are going to map the states of all the cells in my brain and body and after a few months they will have constructed a perfect duplicate of me out of organic matter. And this duplicate will wake up fully psychologically continuous with me, seeming to remember my life with my character, etc.

VESEY Yes.

PARFIT Now, in this case, which is a secular version of the Resurrection, we're very inclined to think that the following question arises and is very real and very important. The question is: 'Will that person who wakes up in three months be *me* or will he be some quite other person who's merely artificially made to be exactly like me?'

VESEY It does seem to be a real question. I mean in the one case, if it's going to be me, then I have expectations and so on, and in the other case, where it isn't me, I don't.

PARFIT I agree, it seems as if there couldn't be a bigger difference between it being me and it being someone else.

VESEY But you want to say that the two possibilities are in fact the same?

PARFIT I want to say that those two descriptions, 'It's going to be me' and 'It's going to be someone who is merely exactly like me', they don't describe different outcomes, different courses of events, only one of which can happen. There are two ways of describing one and the same course of events. What I mean by that perhaps could be shown if we take an exactly comparable case involving not a person but something about which I think we're not inclined to have a false view.

VESEY Yes.

PARFIT Something like a club. Suppose there's some club in the nineteenth century ...

VESEY The Sherlock Holmes Club or something like that?

PARFIT Yes, perhaps. And after several years of meeting it ceases to meet. The club dies.

VESEY Right.

PARFIT And then two of its members, let's say, have emigrated to America, and after about fifteen years they get together and they start up a club. It has exactly the same rules, completely new membership except for the first two people, and they give it the same name. Now suppose someone came along and said: 'There's a real mystery here, because the following question is one that must have an answer, but how can we answer it?' The question is, 'Have they started up the very same club – is it the same club as the one they belonged to in England? – or is it a completely new club that is just exactly similar?'

VESEY Yes.

PARFIT Well, in that case we all think that this man's remark is absurd; there's no difference at all. Now that's my model for the true view about the case where they make a duplicate of me. It seems that there's all the difference in the world between its being me and it's being this other person who's exactly like me. But if we think there's no difference at all in the case of the clubs, why do we think there's a difference in the case of personal identity, and how can we defend the view that there's a difference?

VESEY Yes, I can see how some people would defend it. I mean, a **dualist** would defend it in terms of a soul being a simple thing, but ...

PARFIT Well, let me try another case which I think helps to ease us out of this belief we're very strongly inclined to hold.

VESEY Go on.

PARFIT Well, this isn't a single case, this is a whole range of cases, a whole smooth spectrum of different cases, which are all very similar to the next one in the range. At the start of this range of cases you suppose that the scientists are going to replace one percent of the cells in your brain and body with exact duplicates.

VESEY Yes.

PARFIT Now if that were to be done, no one has any doubt that they'd survive. I think that's obvious because after all you can lose one percent of the cells and survive. As we get further along the range they replace a larger and larger percentage of cells with exact duplicates, and of course at the far end of this range, where they replace a hundred percent, then we've got my case where they just make a duplicate out of wholly fresh matter.

VESEY Yes.

PARFIT Now on the view that there's all the difference in the world between it being me and it being this other person who is exactly like me, we ought in consistency to think that in some case in the middle of that range, where, say, they're going to replace fifty percent, the same question arises: is it going to be *me* or this completely different character? I think that even the most convinced dualist who believes in the soul is going to find this range of cases very embarrassing, because he seems committed to the view that there's some crucial percentage up to which it's going to be him and after which it suddenly ceases to be him. But I find that wholly unbelievable.

VESEY Yes. He's going to have to invent some sort of theory about the relation of mind and body to get around this one. I'm not quite sure how he would do it. But, Derek, could we go on to a related question? Suppose I accepted what you said, that is, that there isn't anything more to identity than what you call psychological continuity in a one-one case. Suppose I accepted that, then I would want to go on and ask you, well, what's the philosophical importance of this?

PARFIT The philosophical importance is, I think, that psychological continuity is obviously, when we think about it, a matter of degree. So long as we think that identity is this further fact, one of the things we're inclined to think is that it's all or nothing, as you said earlier. Well, if we give up that belief and if we realize that what matters in my continued existence is a matter of degree, then this does make a difference in actual cases. All the cases that I've considered so far are of course bizarre science-fiction cases. But I think that in actual life it's obvious on reflection that, to give an example, the relations between me now and me next year are much closer in every way than the relations between me now and me in twenty years. And the sorts of relations that I'm thinking of are relations of memory, character, ambition, intention – all of those. Next year I shall remember much more of this year than I will in twenty years. I shall have a much more similar character. I shall be carrying out more of the same plans, ambitions and, if that is so, I think there are various plausible implications for our moral beliefs and various possible effects on our emotions.

VESEY For our moral beliefs – what have you in mind?

PARFIT Well, let's take one very simple example. On the view which I'm sketching it seems to me much more plausible to claim that people deserve much less punishment, or even perhaps no punishment, for what they did many years ago as compared with what they did very recently. Plausible because the relations between them now and them many years ago when they committed the crime are so much weaker.

VESEY But they are still the person who are responsible for the crime.

PARFIT Well ... I think you say that because even if they've changed in many ways, after all it was just as much them who committed the crime. I think that's true, but on the view for which I'm sketching, we would come to think that it's a completely trivial truth. It's like the following truth: it's like the truth that all of my relatives are just as much my relatives. Suppose I in my will left more money to my close relatives and less to my distant relatives; a mere pittance to my second cousin twenty-nine times removed. If you said, 'But that's clearly unreasonable because all of your relatives are just as much your relatives', there's a sense in which that's true but it's obviously too trivial to make my will an unreasonable will. And that's because what's involved in kinship is a matter of degree.

VESEY Yes.

PARFIT Now, if we think that what's involved in its being the same person now as the person who's committed the crime is a matter of degree, then the truth that it was just as much him who committed the crime, will seem to us trivial in the way that the truth 'all my relatives are equally my relatives' is obviously trivial.

VESEY Yes. So you think that I should regard myself in twenty years' time as like a fairly distant relative of myself?

PARFIT Well, I don't want to exaggerate; I think the connections are much closer.

VESEY Suppose I said that this point about psychological continuity being a matter of degree – suppose I said that this isn't anything that anybody denies?

PARFIT I don't think anybody does on reflection deny that psychological continuity is a matter of degree. I agree. But I think what they may deny, and I think what may make a difference to their view, if they come over to the view for which I'm arguing – what they may deny is that psychological

continuity is all there is to identity. Because what I'm arguing against is this further belief that I think we're all inclined to hold even if we don't realise it: the belief that however much we change, there's a profound sense in which the changed us is going to be just as much us, that even if some magic wand turned me into a completely different sort of person – a prince with a totally different character, mental powers – it would be just as much me. That's what I'm denying.

VESEY Yes. This is the belief which I began by stating, and I think that if we did lose that belief that would be a change indeed.

## Notes

- 1 *Sunday Times*, 9 December, 1973, p. 13.
- 2 Shoemaker, S. (1963) *Self-Knowledge and Self-Identity*, Ithaca, NY, Cornell University Press, pp. 23–4.
- 3 Parfit, D. (1971) 'On the importance of self identity', *The Journal of Philosophy*, vol. 68, no. 20, p. 15.

# Glossary

## bundle theory of the self

David Hume's view that the illusion we experience of having a self is the result of our having a succession of perceptions bundled together by memory.

## dualist

A mind/body dualist is someone who believes that mind and body are distinct.

## empiricism

This is the view shared by John Locke and David Hume and many other philosophers that experience is the source of our knowledge of the world. Empiricism is often contrasted with Rationalism. Rationalist philosophers, such as René Descartes, believed that we could acquire knowledge by the use of reason and that the senses were unreliable in this respect. The traditional division between Empiricists and Rationalists, however, is not as clear cut as it is often claimed to be.

## forensic

John Locke uses the word 'forensic' to refer to questions of justifiable blame and punishment for individual actions. He describes 'person' as a forensic term in this sense.

## idea

For David Hume, an idea is a copy of an impression. See **impression** below.

## identity

*See numerically identical, qualitatively identical.*

## impression

A technical term that David Hume used to refer to any direct sensory input. So, for example, when you look at the sky you have an impression of it while you are looking at it. This contrasts with ideas that are, according to Hume, copies of impressions. When you close your eyes and remember the sky, you are having an idea of the sky, not an impression.

**man**

For John Locke, the word ‘man’ contrasts with the word ‘person’. ‘Man’ means more or less what we would mean by ‘human being’. You are the same man you were if there is bodily continuity of the appropriate kind between then and now. But you can be the same man without being the same person, for Locke. He thought that continuity of memory determined whether or not you are the same person that you were.

**numerically identical**

If two things are numerically identical, that means they are the same thing. The Morning Star and the Evening Star are numerically identical: that is to say, they are one and the same thing (the Morning Star and the Evening Star are two names used to refer to the planet Venus). See also **qualitatively identical**.

**person**

For John Locke, ‘person’ is a morally significant term. You are the same person you were formerly if you have continuity of memory with that person. This makes you morally responsible for the actions you committed earlier. Locke believed that you could be the same **man** that you were, i.e. the same physical being (despite some bodily changes), but yet you might not be the same person. The question of whether or not you were the same person turned for Locke on whether or not you could remember what you did earlier. The question of whether you were the same man or not turned on whether you had bodily continuity of the appropriate sort.

**qualitatively identical**

If two things are qualitatively identical, that means they are of the same type. So, for example, identical twins have qualitatively identical chromosomes, but are still numerically distinct: that is, there are two of them. See also **numerically identical**.

**reductionism about personal identity**

In the context of Derek Parfit’s account of the self, reductionism is the view that what makes someone the same person over time is nothing more than some physical and psychological facts. Facts about bodies and brains and their continuity answer all questions about personal identity. This stands in contrast with non-reductionism about personal identity, such as the view that a person is an immaterial soul.

**soul**

A term that religious people and others use to refer to an allegedly immaterial aspect of their being that may be able to exist independently of their body.

**thought experiment**

An imagined situation that brings out our intuitions about a particular case. Thought experiments allow philosophers to think about hypothetical situations, controlling variables.

# Acknowledgements

Grateful acknowledgement is made to the following sources:

## Cover

René Magritte, *La Reproduction*, 1937, oil on canvas, 81 x 65 cm. Museum Boijmans van Beuningen, Rotterdam. © ADAGP, Paris and DACS, London 2010. Photo: © Scala, Florence. Frame: © istockphoto.com/subjug.

## Reading 1

© Jonathan Bennett, 'Some texts from early modern philosophy', [www.earlymoderntexts.com](http://www.earlymoderntexts.com)

## Reading 2

© Problems From Locke by Mackie 1778w from pp. 173–7. By permission of Oxford University Press.

# Index

Page numbers in **bold** refer to figures.

- activities 11
- afterlife, belief in 8–9
- ageing and personal identity
  - bodily change and continuity 8, 9–10
  - and moral responsibility 17
  - Parfit on 88
- Alzheimer's disease, personal identity and moral responsibility 17
- ancient Greek philosophers 33
- audio recordings v
  - 'Hume: an introduction' 58, 61
  - 'Hume on the self' 58, 64, 71–2
  - 'Locke on persons' 32, 43, 46, 47
  - 'Parfit: an overview' 83–4, 85, 86, 98
  - 'Parfit on personal identity' 85–6, 115–25
  - 'Personal identity' 18–19, 24, 43, 100
  - 'Reading historical texts' 32, 50
  - 'Thought experiments' 14–16, 24, 92–3
- Baggini, Julian
  - audio interview on thought experiments 14–16, 24, 92–3
  - The Pig That Wants to be Eaten* 14
- Beebee, Helen, audio interview on Parfit 83–4, 85, 86, 98
- Bennett, Jonathan, paraphrase of Locke on personal identity 32, 38, 107–9
- Bhagavad Gita* 8–9
- Blackburn, Simon 61, 71–2
  - How to Read Hume 64
- bodily continuity and personal identity 7–10, 13, 19
  - in Locke 35–6, 39–40, 45, 59, 91
  - in Parfit 90–1, 95–6
- Boyle, Robert 59
- the brain and personal identity 9, 18
  - multiple personalities 22
  - Parfit on 85–6, 115–25
- Buddhism
  - and Parfit on personal identity 91
  - and reincarnation 8
- Butler, Joseph 71, 100
  - The Analogy of Religion* 66
  - Of Personal Identity* 65

- children, physical and character changes in 7–8
- consciousness
  - and personal identity in Locke 39–41, 51, 59, 99
  - Mackie on 46–7, 110–14
- conversation, philosophy as ongoing 48–50
- criminal behaviour, personal identity and moral responsibility 17
- critical thinking vi
- Cutty Sark* restoration 11–12, 13
- Day of Judgment
  - and bodily resurrection 36, 59
  - and Locke on moral responsibility 36–7, 42, 47
- death
  - and ageing 8, 9
  - bodily resurrection after 8, 59, 71
  - and Hume's concept of the self 71
  - Parfit and the significance of 88–91, 95–6
  - and Parfit's thought experiments 87
  - possibility of existence after 8–9
- Dementrius Phalerius 11, 12
- Descartes, René, *Discourse on Method* 48–9
- Fricker, Miranda 50
- God
  - Hume's attack on arguments for God's existence 62
  - and Locke on personal identity 36–7, 42, 43, 47
- Grayling, A.C., audio interviews 18–19, 24, 43, 47, 61, 64, 100
- Hobbes, Thomas
  - De Corpore* 14
  - Leviathan* 14
  - and the Ship of Theseus 14
- Hume, David v, 10, 18, 19, 50, 51, 55–75, 62
  - audio recordings
    - 'Hume: an introduction' 58, 61
    - 'Hume on the self' 58, 64, 71–2
  - biographical details 61–2
  - on change, continuity and identity 72–3
  - Dialogues concerning Natural Religion* 61, 62
  - Enquiries concerning Human Understanding* 63–4, 67, 68–70
  - A History of England* 61

Hume's Fork 69–70  
 on impressions and ideas 61, 66–70, 74  
 and Mackie 46  
 and Parfit 75, 83  
 reading 33–4  
 and the self as fiction 73  
*A Treatise of Human Nature* 61  
 'Of the origin of our ideas' 67–8  
 on the self as bundle 58, 59, 62, 63–4, 65–6

ideas  
 Hume on impressions and ideas 61, 66–70, 74  
 philosophy as a history of 48, 50

identity  
 Hume on change, continuity and identity 72–3  
 Locke and the identity of consciousness 40–1  
 impressions, Hume on impressions and ideas 61, 66–70, 74  
 Iraq war 36

James, Susan 50  
 Johnson, W.J. 9  
 Jones, Peter 61

Kafka, Franz, 'Metamorphosis' 21  
 Kant, Immanuel 72, 74  
*Critique of Pure Reason* 35

Lochner, Stephan, *The Last Judgement* 37  
 Locke, John v, 10, 12, 18–19, 20, 27–47  
 audio recording, 'Locke on persons' 32, 43, 46, 47  
 Bennett's paraphrase of Locke on personal identity 32, 38, 107–9  
 biographical information 20  
*Essay concerning Human Understanding* 10, 20  
 and Hume 59, 61–2, 72  
 Mackie's reading of 46–7, 110–14  
 memory and personal identity in 25, 39–40, 41–2, 45–6  
 on moral responsibility 33, 36–7, 41–2, 46–7, 51, 99  
 and Parfit 75, 83, 90  
 reading as a historical text 35–8, 48  
 Reid's criticism of 43–6, 51, 59, 91, 99–100  
 terminology 35–8  
 thought experiments 19–24, 33, 92, 93, 99  
 Little Finger 39, 39–40  
 'The day-man and the night-man' 21–2, 23, 93, 99

'The drunk and the sober man' 22, 23–4  
 'The prince and the cobbler' 20–1, 23, 91, 93, 99

Mackie, J.L., *Problems from Locke*, 46–7, 110–14  
 madness and personal identity 41–2  
 man, Locke's use of the term 35–6  
 memory and personal identity 17, 18, 19, 25  
 gradual memory loss 43–6  
 and Hume's 'Bundle Theory of the Self' 63  
 in Locke 25, 39–40, 41–2, 45–6  
 in Parfit  
 duplicate-body case 87  
 teletransporter case 90–1  
 thought experiments on 19–24, 39–40

Millican, Peter 61

moral responsibility  
 and personal identity 17, 19, 22–4  
 Locke on 33, 36–7, 41–2, 46–7, 51, 99  
 Mackie on 46–7

multiple personalities 21–2, 23, 93

numerical identity 12

objects changed over time, continuing identity of 11–14, 15–16, 72–3  
*Oxford Dictionary of Philosophy* 24

paintings, Rembrandt's self-portraits 7  
 Parfit, Derek v, 10, 18, 19, 47, 51, 64, 81–98, 100  
 audio recordings  
 'Parfit: an overview' 83–4, 85, 86, 98  
 'Parfit on personal identity' 85–6, 115–25  
 and Hume 75, 83  
 and Locke 75, 83, 90  
 on personal identity 115–25  
 reading 33–4  
*Reasons and Persons* 82, 88, 93  
 rejection of the reductionist view 94  
 and the significance of death 88–91, 95–6  
 successive selves idea 91–2  
 Taylor's criticism of 97–8  
 thought experiments 75  
 relevance of 92–4  
 the teletransporter 89–91

person, Locke's use of the term 36–7  
 personal identity 17–25  
 and bodily continuity 7–10, 13, 19

in Locke 35–6, 39–40, 45, 59  
 and the brain 9, 18, 22  
 Grayling on 18–19  
 Hume and the self as fiction 18, 19, 61–75  
 and moral responsibility 17, 19, 22–4  
     Locke on 33, 36–7, 41–2, 46–7, 51  
     Mackie on 46–7  
 and multiple personalities 21–2, 23, 93  
 Parfit on the self and the future 83–100  
 philosophical meaning of 18  
 thought experiments on 19–24  
 plants, identity of 13, 19  
 Plato 49  
 Plutarch, and the Ship of Theseus 11  
 psychological continuity  
     in Locke 39–41, 59, 99  
     Reid's criticism of 99–100  
     Parfit on 84, 85–6, 91, 95–6  
     Taylor's criticism of 97–8  
 punishment  
     and bodily resurrection 59  
     personal identity and moral responsibility 17, 22–4  
     Locke on 41–2, 46–7, 51  
 qualitative identity 12  
 Quine, W.V.O., criticism of thought experiments 92, 93  
 Ramsay, Allan, *David Hume* 62  
 reading historical texts 33–4, 48–50  
     audio recording 32, 50  
     as history of ideas 48, 50  
     Locke 35–8, 48  
     as part of a conversation 48–50  
     translations of 35  
 real-life cases, and thought experiments 92  
 Reid, Thomas 31, 44  
     criticism of Locke 43–6, 51, 59, 91, 99–100  
     *An Inquiry into the Human Mind, on the Principles of Common Sense* 43  
 reincarnation 8–9, 38, 71  
 religion  
     and bodily resurrection after death 8, 59, 71  
     and reincarnation 8–9, 38, 71  
 Rembrandt van Rijn, self-portraits by 7  
 Saddam Hussein 36  
 Schopenhauer, Arthur, 'On thinking for oneself' 49–50  
 scientific experiments, and thought experiments 14, 15  
*Seven Up!* (Granada documentary television series) 7–8  
 Ship of Theseus case 11–16  
     audio recording on 14–16  
     Hobbes's version of 14  
 Snowdon, Paul 90  
     'Locke on persons' (audio interview) 43, 46, 47  
 the soul  
     and Hume's search for the self 63, 69, 71, 73, 100  
     in Locke 37–8, 59  
     and Parfit on personal identity 87  
     and reincarnation 38  
 Spencer, Sir Stanley, *The Resurrection, Cookham* 9  
*Star Trek*, teletransporter in 89–91  
 substance  
     use of the term in Locke 37–8  
     *see also* the soul  
 survival, Parfit on personal identity and 86  
 Taylor, Charles, *Sources of the Self* 82, 97–8, 100  
 Theseus, Ship of 11–16  
 thought experiments v, 10, 19–24, 25, 33  
     Locke 19–24, 33, 92, 93, 99  
     Little Finger 39, 39–40  
     'The day-man and the night-man' 21–2, 23, 93, 99  
     'The drunk and the sober man' 22, 23–4  
     'The prince and the cobbler' 20–1, 23, 91, 93, 99  
 Parfit 75  
     brain bisection 85–6  
     the duplicate that survives death 87  
     the teletransporter 89–91  
     relevance of 92–4  
     and scientific experiments 14, 15  
     Ship of Theseus 11–16  
     audio recording on 14–16  
     trolley problem 15  
     transitivity of identity 45  
     trolley problem thought experiments 15  
 Vesey, Godfrey 85  
 Whitehead, Alfred 49  
 Wilkes, Kathleen 92

A222 Exploring philosophy

# The Self

Nigel Warburton